



When did the ancestor of true bugs become stinky?

Disentangling the phylogenomics of Hemiptera–Heteroptera

Wang, Yan Hui; Wu, Hao Yang; Rédei, Dávid; Xie, Qiang; Chen, Yan; Chen, Ping Ping; Dong, Zhuo Er; Dang, Kai; Damgaard, Jakob; Štys, Pavel; Wu, Yan Zhuo; Luo, Jiu Yang; Sun, Xiao Ya; Hartung, Viktor; Kuechler, Stefan M.; Liu, Yang; Liu, Hua Xi; Bu, Wen Jun

Published in:
Cladistics

DOI:
[10.1111/cla.12232](https://doi.org/10.1111/cla.12232)


Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
[CC BY](#)

Citation for published version (APA):
Wang, Y. H., Wu, H. Y., Rédei, D., Xie, Q., Chen, Y., Chen, P. P., Dong, Z. E., Dang, K., Damgaard, J., Štys, P., Wu, Y. Z., Luo, J. Y., Sun, X. Y., Hartung, V., Kuechler, S. M., Liu, Y., Liu, H. X., & Bu, W. J. (2019). When did the ancestor of true bugs become stinky? Disentangling the phylogenomics of Hemiptera–Heteroptera. *Cladistics*, 35(1), 42-66. <https://doi.org/10.1111/cla.12232>

When did the ancestor of true bugs become stinky? Disentangling the phylogenomics of Hemiptera–Heteroptera

Yan-Hui Wang^{a,b,c,†}, Hao-Yang Wu^{a,b,c,†}, Dávid Rédei^{c,†}, Qiang Xie^{a,b,c,*} , Yan Chen^d, Ping-Ping Chen^e, Zhuo-Er Dong^c, Kai Dang^c, Jakob Damgaard^f, Pavel Štys^g, Yan-Zhuo Wu^c, Jiu-Yang Luo^c, Xiao-Ya Sun^c, Viktor Hartung^{h,i}, Stefan M. Kuechler^j, Yang Liu^c, Hua-Xi Liu^c and Wen-Jun Bu^c

^aDepartment of Ecology and Evolution, College of Life Sciences, Sun Yat-sen University, 135 Xingangxi Road, Guangzhou 510275, Guangdong, China; ^bState Key Laboratory of Biocontrol, Sun Yat-sen University, 135 Xingangxi Road, Guangzhou 510275, Guangdong, China; ^cInstitute of Entomology, College of Life Sciences, Nankai University, 94 Weijin Road, 300071 Tianjin, China; ^dChinese Academy of Inspection and Quarantine, No. A3, Gaobeidian Bei Lu, Chaoyang District, Beijing, 100123, China; ^eNetherlands Centre of Biodiversity Naturalis, 2300 RA, Leiden, Netherlands; ^fNatural History Museum of Denmark, Universitetsparken 15, 2100 Copenhagen Ø, Denmark; ^gDepartment of Zoology, Faculty of Science, Charles University in Prague, Viničná 7, CZ-128 44 Praha 2, Czech Republic; ^hStaatliches Museum für Naturkunde Karlsruhe, Erbprinzenstrasse 13, 76133 Karlsruhe, Germany; ⁱMuseum für Naturkunde — Leibniz-Institute for Research on Evolution and Biodiversity, Invalidenstrasse 43, 10115 Berlin, Germany; ^jDepartment of Animal Ecology II, University of Bayreuth, Universitätsstrasse 30, 95440 Bayreuth, Germany

Accepted 7 November 2017

Abstract

The phylogeny of true bugs (Hemiptera: Heteroptera), one of the most diverse insect groups in terms of morphology and ecology, has been the focus of attention for decades with respect to several deep nodes between the suborders of Hemiptera and the infraorders of Heteroptera. Here, we assembled a phylogenomic data set of 53 taxa and 3102 orthologous genes to investigate the phylogeny of Hemiptera–Heteroptera, and both concatenation and coalescent methods were used. A binode-control approach for data filtering was introduced to reduce the incongruence between different genes, which can improve the performance of phylogenetic reconstruction. Both hypotheses (Coleorrhyncha + Heteroptera) and (Coleorrhyncha + Auchenorrhyncha) received support from various analyses, in which the former is more consistent with the morphological evidence. Based on a divergence time estimation performed on genes with a strong phylogenetic signal, the origin of true bugs was dated to 290–268 Ma in the Permian, the time in Earth's history with the highest concentration of atmospheric oxygen. During this time interval, at least 1007 apomorphic amino acids were retained in the common ancestor of the extant true bugs. These molecular apomorphies are located in 553 orthologous genes, which suggests the common ancestor of the extant true bugs may have experienced large-scale evolution at the genome level.

© 2017 The Authors. *Cladistics* published by John Wiley & Sons Ltd on behalf of Willi Hennig Society.

Introduction

Hemiptera is the largest order of non-holometabolous insects, with ~100 000 described species belonging to ~180 families (Footitt and Adler, 2009). Five basic

monophyletic groups have repeatedly been recognized within the group: Sternorrhyncha (scale insects, aphids, whiteflies and psyllids), Fulgoromorpha (planthoppers), Cicadomorpha (leafhoppers, spittlebugs and cicadas), Coleorrhyncha (moss bugs) and Heteroptera (true bugs). While there is broad agreement regarding the phylogenetic position of Sternorrhyncha as the sister group of the remaining four clades (= Euhemiptera), the relationships within Euhemiptera

*Corresponding author.

E-mail address: xieq8@mail.sysu.edu.cn

[†]These authors contributed equally to this work.

remain contentious (Fig. 1). Fulgoromorpha and Cicadomorpha have been treated either as two separate suborders (Campbell et al., 1995; von Dohlen and Moran, 1995; Sorensen et al., 1995; Xie et al., 2008; Song et al., 2012, 2016; Cui et al., 2013) or united into a single suborder Auchenorrhyncha (Cryan and Urban, 2012; Friedemann et al., 2014; Li et al., 2017). Additionally, either Coleorrhyncha (Campbell et al., 1995; Sorensen et al., 1995; Ouvrard et al., 2000; Xie et al., 2008; Cryan and Urban, 2012; Li et al., 2015, 2017), Cicadomorpha (Cui et al., 2013), Fulgoromorpha (Goodchild, 1966; Bourgoin, 1988, 1993; Campbell et al., 1994; von Dohlen and Moran, 1995), combined Coleorrhyncha and Fulgoromorpha (Song et al., 2016) or even all of the three lineages mentioned above as a whole (Misof et al., 2014) have been suggested to be the sister group of Heteroptera. Among these five competing hypotheses, the sister group relationship between Coleorrhyncha and Heteroptera has received the strongest support from molecular (Wheeler et al., 1993; Campbell et al., 1995; Sorensen et al., 1995; Ouvrard et al., 2000; Xie et al., 2008; Cryan and Urban, 2012; Li et al., 2015, 2017), morphological (Spangenberg et al., 2013; Friedemann et al., 2014) and cytogenetical data (Kuznetsova et al.,

2015). The clade Coleorrhyncha + Heteroptera has been named Heteropteroidea (Schlee, 1969), Heteropteroidea (Zrzavý, 1992) or Prosorrhyncha (Sorensen et al., 1995).

The oldest definitive fossil of a hemipteran, *Aviorrhyncha magnifica*, can be dated to ~310 Ma, the Moscovian stage of the Carboniferous (Nel et al., 2013). During their more than 300-Myr evolutionary history, hemipterans have diversified in both their morphology and their ecology. Heteroptera is the most diverse clade of Hemiptera in terms of habitat and life habit (Henry, 2009), including groups colonizing all major habitat types and comprising phytophagous, zoophagous and ectoparasitic species. Correspondingly, the taxonomy of Heteroptera is also complex. Heteroptera is universally regarded as a monophyletic group characterized by several synapomorphies, including the presence of paired metathoracic scent glands in adults and dorsal abdominal scent glands in nymphs, and a prognathous head with distinctly developed gula (Slater, 1982; Schuh and Slater, 1995). The suborder is subdivided into seven infraorders (Štys and Kerzhner, 1975) and 24 superfamilies (Schuh and Slater, 1995; Henry, 2009). Despite wide recognition of the infraorders, their evolutionary relationships have

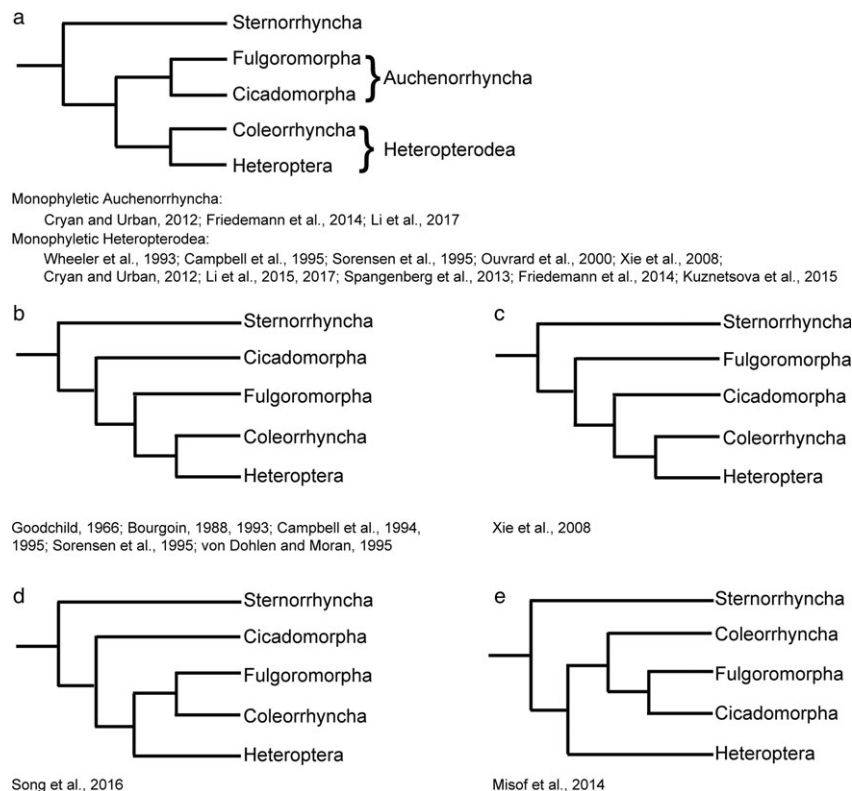


Fig. 1. Alternative hypotheses of higher-level relationships within Hemiptera. (a) monophyletic Auchenorrhyncha as sister group to Heteropteroidea; (b) nonmonophyletic Auchenorrhyncha, with Fulgoromorpha as sister group to Heteropteroidea; (c) nonmonophyletic Auchenorrhyncha, with Cicadomorpha as sister group to Heteropteroidea; (d) nonmonophyletic Auchenorrhyncha and Heteropteroidea, with Fulgoromorpha as sister group to Coleorrhyncha; and (e) monophyletic Auchenorrhyncha as sister group to Coleorrhyncha.

remained uncertain, and at least six different phylogenetic hypotheses have been raised in the past four decades (Schuh, 1979; Wheeler et al., 1993; Shcherbakov and Popov, 2002; Xie et al., 2008; Weirauch and Štys, 2014; Wang et al., 2016). The only congruencies between the various hypotheses are the monophyly of Terheteroptera (= Cimicomorpha + Pentatomomorpha) and the unnamed clade formed by Lepidodromomorpha + Terheteroptera, while the phylogenetic positions of the other four infraorders (i.e., Dipsocoromorpha, Enicocephalomorpha, Gerromorpha and Nepomorpha) vary across analyses.

In addition to their complicated diverse habitats, life habits and classification systems, several true bug species have received intense focus for economic, medical or scientific reasons. These include, for example, diverse members of pentatomid bugs that act as globally severe agricultural pests, e.g., the brown marmorated stink bug (Pentatomidae: *Halyomorpha halys*) (Garipey et al., 2014), the common bed bug (Cimicidae: *Cimex lectularius*) as a resurgent worldwide public health pest (Doggett et al., 2012), the kissing bugs (Reduviidae: Triatominae) as vectors of Chagas disease (De Noya et al., 2010), the milkweed bug (Lygaeidae: *Oncopeltus fasciatus*) as a model organism widely used in Evo-Devo studies, and the water strider (Gerridae: *Aquarius remigis*) as a model for biomechanical studies of hydrodynamics (Hu et al., 2003). However, until now, the time of origin of the Heteroptera has remained obscure.

In the genomic era, phylogeneticists can simultaneously utilize hundreds or thousands of single-copy nuclear genes to infer evolutionary relationships (Lemmon and Lemmon, 2013). However, as the quantity of phylogenomic data increases, only stochastic errors caused by limited data sampling are effectively alleviated (Rokas et al., 2003; Delsuc et al., 2005), while the impact of systematic bias, such as the heterogeneity in evolutionary rates among lineages (Irisarri and Meyer, 2016), has not yet been effectively eliminated. Therefore, merely increasing the number of gene sequences does not necessarily resolve phylogenetic incongruences, and selecting data with strong phylogenetic signal is of crucial importance (Philippe and Roure, 2011).

Several optimization approaches have been suggested to produce subsets of data with increased phylogenetic signal. Such methods can be divided into four broad types. First, genes with broader taxon coverage and less missing data are given preference; rogue taxa and poorly aligned genes are removed, or the amount of missing data is decreased (Meusemann et al., 2010; Dell'Ampio et al., 2014). Second, genes with certain traits, such as stationary base composition (Collins et al., 2005; Romiguier et al., 2013) or low evolutionary rates (Regier et al., 2008; Philippe et al.,

2009; Zhang et al., 2012; Betancur-R et al., 2014), are given preference. Third, preference is given to genes that perform well in recovering the correct topology, which is congruent with other lines of evidence. This process is achieved by filtering genes with high resolution based on the average bootstrap values of their resulting gene trees (Salichos and Rokas, 2013; Chen et al., 2015). The fourth approach is different from the above three in that it does not aim to optimize data sets as a whole. It instead focuses on particular phylogenetic questions and optimizes data with the method of node control, which is implemented by filtering genes based on their corresponding bootstrap support values along with the monophyly of a question-specific bipartition (Salichos and Rokas, 2013; Chen et al., 2015). Among all of these approaches, the method of node control concentrates more on the target question and was shown to be the most effective (Salichos and Rokas, 2013) or one of the most effective two in gene filtering or optimizing phylogenomic data sets (Chen et al., 2015). In this study, this approach was optimized to use one or two question-specific clades with relatively high support values as controlled nodes to filter genes, and term this modified method “uninode control” or “binode control”.

In the present study, we assembled a data set of 3102 orthologous genes from 53 taxa (51 ingroups covering all major lineages of Hemiptera and two outgroup thrips) for resolving ancient nodes in the phylogeny of Hemiptera–Heteroptera. Our results found that the phylogenetic results inferred from the genes filtered by binode control or uninode control are more congruent with other lines of evidence. The divergence times of the major subclades of Hemiptera were then estimated based on the filtered genes with strong phylogenetic signal. The common ancestor of true bugs was found to originate between 290 and 268 Ma in the Permian, the time in Earth's history with the highest concentration of oxygen in the atmosphere.

Materials and methods

Taxon sampling

Our taxon sampling included 53 species, of which 51 species of Hemiptera were ingroups and two species of Thysanoptera were outgroups (Table 1). The sampled species covered all five potential suborders of Hemiptera and the seven infraorders of Heteroptera. Thirty-two hemipteran transcriptomes were *de novo* sequenced in this study, including two infraorders and 21 families of Heteroptera that were analysed based on transcriptomic data for the first time. Genome assemblies of *Acyrthosiphon pisum* were downloaded from the

Ensembl Metazoa website (<http://metazoa.ensembl.org>). Orthologous protein-coding genes of *Frankliniella occidentalis*, *Gerris buenoi*, *Rhodnius prolixus*, *Cimex lectularius* and *Oncopeltus fasciatus* were downloaded from OrthoDB (<http://www.orthodb.org/>). Transcriptome assemblies of the remaining species were downloaded from the NCBI Transcriptome Shotgun Assembly (TSA) sequence database (<http://www.ncbi.nlm.nih.gov/Traces/wgs/>). The complete list of taxa and the corresponding sources used in this study are provided in Table 1.

Transcriptome assembly and orthology assignment

Total RNA of each species (adults) was extracted using TRIzol reagent according to the manufacturer's instructions (Invitrogen, Grand Island, NY, USA). RNA quality parameters, such as quantification and RNA integrity number (RIN), were assessed using an Agilent 2100 Bioanalyzer. All samples had RNA integrity values above 6.5. Individual libraries were prepared using the Dynabeads mRNA Purification Kit followed by transcribing of mRNA using various enzymes (Invitrogen), according to the manufacturer's instructions. The cDNA library was subsequently sequenced on an Illumina HiSeq2000/4000 device, according to the manufacturer's protocols. We obtained ~4 billion bases of raw data for each of the 32 species after sequencing. The raw sequence reads have been deposited at the NCBI Sequence Read Archive (SRA) (BioProject ID: PRJNA352589).

Raw data were pre-processed by removing reads of poor quality. The remaining data were used for *de novo* transcriptome assembly. After reads filtering, Trinity was used to conduct *de novo* assembly of each species (Haas et al., 2013). To obtain non-redundant transcript sequences, contigs of each species were further clustered to unigenes using the TGI Clustering tool (Pertea et al., 2003). Orthologous transcript prediction was performed using HaMStR.v13.2.6 (Ebersberger et al., 2009). As HaMStR needs to map candidate transcripts to core-orthologue genes that are known to occur in single-copy in the sequenced genomes of reference species, we established our own core-orthologue genes and local databases of the protein sequences of three reference species, *Cimex lectularius*, *Oncopeltus fasciatus* and *Rhodnius prolixus*. The detailed protocols of establishing core-orthologue genes and local databases are provided in File S1. Altogether, we established 4069 core-orthologue groups. For each taxon, orthologous gene prediction was performed based on the 4069 core-orthologue groups and the three reference species mentioned above. In HaMStR, parameters were set as follows: the e-value cut-off for pHMM search and BLAST was set as 1E-5; the presence of orthology depends on whether best reciprocal hits (BRH) of BLAST can be

found between candidate transcript sequences and at least one of the three reference species (option: *relaxed*); the single best or the best set of non-overlapping transcripts was kept in the case that multiple transcripts had been assigned to a given orthologous gene (OG) (option: *representative*). Non-overlapping transcripts were automatically concatenated.

Alignment and alignment masking

Compared with DNA sequences, amino acid sequences are thought to be less subject to systematic errors such as long-branch attraction or compositional variation among distantly related lineages (Simmons et al., 2004; Lartillot et al., 2007). Therefore, in this study, all phylogenetic analyses were based on amino acid sequences. After orthology prediction for each species, Phylopipe.v1.1 (<http://sourceforge.net/projects/phylopipe/>) was used to combine orthologous genes present in ≥50% of taxa (option: *cmb_factor* = 50). The amino acid sequences of each OG were preliminarily aligned using Mafft.v7.222 (Katoh and Standley, 2013) with the alignment algorithm L-INS-i. Subsequently, the Mafft alignments were optimized with Muscle.v3.8.31 (Edgar, 2004) using the refine option. The batch processing of the alignment for each OG was achieved using an in-house script. To remove ambiguously aligned or highly divergent regions in the alignments, Aliscore.v2.2 (Misof and Misof, 2009; Kück et al., 2010) and Alicut (<https://www.zfmk.de/en/research/research-centres-and-groups/utilities>) were used. The software Aliscore was used to identify blocks of putative alignment ambiguities or randomly aligned sections of each OG separately, while Alicut was used to cull the aligned sites with low scores. Sequences with 100% gaps and alignments shorter than 35 amino acids were discarded. Finally, a data set of 3102 orthologous genes was achieved.

Phylogenetic analyses of the concatenated 3102 genes

For the 3102 genes, we constructed the phylogenetic tree using three different methods, maximum likelihood (ML), maximum parsimony (MP) and coalescent-based tree-inference. Both ML and MP analyses are based on the concatenated supermatrix of the 3102 orthologous genes. ML analysis was conducted with RAxML.v8.2.8 using the version Pthreads (Stamatakis, 2014). The best-fitting model was determined by PartitionFinder.v1.1.1 (Lanfear et al., 2012) according to the two-step method provided by Misof et al. (2014). A gamma model of rate heterogeneity was used to account for among-site rate variation. The 3102 genes were grouped into 1395 clusters by PartitionFinder, in which the best fitting model for 1080 clusters was LG. Node support was assessed with 200 replicates using a

Table 1
Taxon sampling used in this study

Order: suborder	Infraorder	Superfamily: family	Species	Data source	Accession number	Number of genes	Number of amino acids
Thysanoptera:		Thripidae	<i>Frankliniella occidentalis</i>	http://www.orthodb.org/	/	2858	722 309
Terebrantia		Phlaeothripidae:	<i>Gynaikothrips ficorum</i>	NCBI TSA	GAXG000000000.2	2529	647 913
Thysanoptera:		Phlaeothripidae	<i>Acyrtosiphon pisum</i>	http://metazoa.ensembl.org/	/	2719	732 551
Tubulifera		Aphidoidea: Aphididae	<i>Bemisia tabaci</i>	NCBI TSA	GAUC000000000.1	2670	672 914
Hemiptera:		Aleyrodoidea: Aleyrodidae	<i>Planococcus citri</i>	NCBI TSA	GAXF000000000.2	2587	715 524
Sternorrhyncha		Coccoidea: Pseudococcidae	<i>Pachypsylla venusta</i>	NCBI TSA	GAOP000000000.1	2711	719 055
		Psylloidea: Psyllidae	<i>Adelges tsugae</i>	NCBI TSA	GBJX000000000.1	2209	730 907
		Phylloxeroidea: Adelgidae	<i>Cercopis vulnerata</i>	NCBI TSA	GAUN000000000.2	2798	702 578
Hemiptera:	Cicadomorpha	Cercopoidea: Cercopidae	<i>Okatanaga villosa</i>	NCBI TSA	GAWQ000000000.2	2658	699 541
Auchenorrhyncha	Cicadomorpha	Cicadoidea: Cicadidae	<i>Graminella nigrifrons</i>	NCBI TSA	GAQX000000000.1	2870	722 106
	Cicadomorpha	Membracoidea: Cicadellidae	<i>Nilaparvata lugens</i>	NCBI TSA	GAXF000000000.2	2861	714 356
	Fulgoromorpha	Delphacidae	<i>Laternaria candelaria</i>	This study	SRX2328724	2360	669 072
	Fulgoromorpha	Flatidae	<i>Geisha distinctissima</i>	This study	SRX2328779	2148	674 088
	Fulgoromorpha	Issidae	<i>Gergithus</i> sp.	This study	SRX2338738	2336	666 262
Hemiptera:	Fulgoromorpha	Peloridiidae	<i>Xenophyes cascus</i>	This study	SRX2328857	2701	690 831
Coleorrhyncha		Peloridiidae	<i>Oiophysa cumberi</i>	This study	SRX2338732	2743	689 965
		Peloridiidae	<i>Xenophysella greensladeae</i>	NCBI TSA	GAXY100000000.2	2475	662 511
Hemiptera:	Dipsozoromorpha	Dipsocoridae	<i>Cryptostemma wygodzinskiyi</i>	This study	SRX2338789	2838	674 523
Heteroptera	Dipsozoromorpha	Schizopteridae	<i>Kokeshia xiei</i>	This study	SRX2338795	2456	627 205
	Dipsozoromorpha	Ceratocombidae	<i>Ceratocombus</i> sp.	This study	SRX2454794	2581	692 697
	Enicocephalomorpha	Aenictopecheidae	<i>Aenictopecheidae</i> sp.	This study	SRX2454792	2837	715 370
	Enicocephalomorpha	Enicocephalidae	<i>Senicopirates</i> sp.	This study	SRX2454790	2566	575 171
	Enicocephalomorpha	Enicocephalidae	<i>Oncolocotis</i> sp.	This study	SRX2454788	2588	685 624
	Gerromorpha	Mesoveloidea: Mesovelidae	<i>Mesovelia thermalis</i>	This study	SRX2455096	2323	616 596
	Gerromorpha	Hebroidea: Hebridae	<i>Hebrus nipponicus</i>	This study	SRX2338908	2803	667 293
	Gerromorpha	Hydrometroidae:	<i>Hydrometra longicapitis</i>	This study	SRX2455000	2263	632 508
		Hydrometridae					
	Gerromorpha	Gerroidea: Gerridae	<i>Gerris buenoi</i>	http://www.orthodb.org/	/	2783	646 277
	Gerromorpha	Gerroidea: Veliidae	<i>Velia caprai</i>	NCBI TSA	GAXO000000000.2	2387	674 885
	Gerromorpha	Gerroidea: Veliidae	<i>Halovelia</i> sp.	This study	SRX2454787	2479	686 690
	Nepomorpha	Nepoidea: Nepidae	<i>Ranatra linearis</i>	NCBI TSA	GAXY200000000.2	2507	654 635
	Nepomorpha	Nepoidea: Belostomatidae	<i>Diplonychus rusticus</i>	This study	SRX2363419	2598	612 603
	Nepomorpha	Ochteroidea: Gelastocoridae	<i>Nerthra indica</i>	This study	SRX2363418	2705	646 441
	Nepomorpha	Corixoidae: Corixidae	<i>Callicorixa gebleri</i>	This study	SRX2454785	2461	664 299
	Nepomorpha	Naucoroidea: Naucoridae	<i>Gestroiella limnocoloides</i>	This study	SRX2454784	2814	648 408
	Nepomorpha	Notonectoidea: Notonectidae	<i>Entithraea bimpressa</i>	This study	SRX2455095	2653	611 976
	Leptopodomorpha	Saldoidea: Saldidae	<i>Calacanthia angulosa</i>	This study	SRX2363506	2703	673 414
	Leptopodomorpha	Saldoidea: Saldidae	<i>Saldula pallipes</i>	This study	SRX2363506	2892	704 319
	Leptopodomorpha	Leptopodoidea: Leptopodidae	<i>Valleriella javanica</i>	This study	SRX2363508	2525	585 559
	Leptopodomorpha	Leptopodoidea: Leptopodidae	<i>Valleriella</i> sp.	This study	SRX2454998	2938	695 313
	Cimicomorpha	Reduvioidae: Reduviidae	<i>Rhodnius prolixus</i>	http://www.orthodb.org/	/	3101	702 166
	Cimicomorpha	Miroidea: Thaumastocoridae	<i>Thaumastocoris</i> sp.	This study	SRX2455003	2723	653 137

Table 1
(Continued)

Order: suborder	Infraorder	Superfamily: family	Species	Data source	Accession number	Number of genes	Number of amino acids
Cimicomomorpha	Cimicomomorpha	Miroidea: Miroidea	<i>Lygus hesperus</i>	NCBI TSA	GBH000000000.1	2892	720 449
Cimicomomorpha	Cimicomomorpha	Miroidea: Tingidae	<i>Dictyla echii</i>	This study	SRX2363509	2839	714 948
Cimicomomorpha	Cimicomomorpha	Naboidea: Nabidae	<i>Alloeohynchus</i> sp.	This study	SRX2455002	2907	676 433
Cimicomomorpha	Cimicomomorpha	Cimicoidea: Cimicidae	<i>Cimex lectularius</i>	http://www.orthodb.org/	/	3102	711 198
Pentatomomorpha	Pentatomomorpha	Aradoidea: Aradidae	<i>Neuroctenus hainanensis</i>	This study	SRX2455094	2327	597 307
Pentatomomorpha	Pentatomomorpha	Pentatomoidea: Acanthosomatidae	<i>Acanthosoma haemorrhoidale</i>	NCBI TSA	GAUV000000000.2	2699	688 991
Pentatomomorpha	Pentatomomorpha	Pentatomoidea: Pentatomidae	<i>Halyomorpha halys</i>	NCBI TSA	GD000000000.1	2795	716 656
Pentatomomorpha	Pentatomomorpha	Pyrhocoroidea: Largidae	<i>Macroheraia grandis</i>	This study	SRX2363511	2512	601 255
Pentatomomorpha	Pentatomomorpha	Pyrhocoroidea: Pyrrhocoridae	<i>Dysdercus cingulatus</i>	This study	SRX2363513	2471	661 270
Pentatomomorpha	Pentatomomorpha	Coreoidea: Coreidae	<i>Homocerus walkeri</i>	This study	SRX2363516	2554	626 735
Pentatomomorpha	Pentatomomorpha	Lygaeoidea: Lygaeidae	<i>Oncopeltus fasciatus</i>	http://www.orthodb.org/	/	3098	629 028
Pentatomomorpha	Pentatomomorpha	Lygaeoidea: Rhyparochromidae	<i>Gyndes</i> sp.	This study	SRX2455001	2322	610 418

rapid bootstrapping algorithm (option: $-f a$). The remaining parameters were left at default settings.

MP analysis was performed using TNT (Goloboff and Catalano, 2016). All characters were equally weighted and gaps were treated as missing characters. Clade robustness was evaluated by using jackknife (JK) resampling (independent character removal) (36%). JK resampling was calculated with 1000 replicates. The traditional searches for trees were performed under TBR (tree bisection reconnection) branch swapping and 20 random replicates.

Gene tree inference and coalescent-based phylogenetic analyses

The unrooted phylogenetic tree of each OG, i.e., gene tree, was inferred using RAXML.v8.2.8 with 150 rapid bootstrap replicates (option: $-f a$) (Stamatakis, 2014). The best-fitting model for each gene was determined by PartitionFinder.v1.1.1 (Lanfear et al., 2012). Through the rapid bootstrap analyses, three files for each OG were obtained, i.e., a best-scoring ML tree, a file containing 150 bootstrap ML trees and a best tree with bootstrap values. The first two files can be applied to infer coalescent-based phylogenetic trees. The last one was used to generate subsets according to different control schemes.

Coalescent-based tree-inference were performed using Astral.v4.10.5 (Mirarab et al., 2014). The phylogenetic tree was inferred from 3102 individual unrooted gene trees and their respective bootstrap replicates using the multi-species coalescent model with 100 bootstrap replicates (options: $-g$ and $-r$). This approach is thought to be more robust to incomplete lineage sorting (ILS) or deep coalescence than concatenation and works quickly on genome-scale data sets (Mirarab et al., 2014).

Subset generation and corresponding phylogenetic analyses

The 3102 best trees with bootstrap values obtained from RAXML were rooted with the thrips (*Frankliniella occidentalis*, *Gynaikothrips ficorum*) using R. These rooted trees were further used for gene filtering. To generate a subset with improved signal quality, we adopted the optimization approach of binode and uninode control to filter signal genes from the 3102 orthologous genes. Three nodes were selected for binode control and uninode control: Euhemiptera, Heteroptera, and Leptopodomorpha and Terheteroptera. For convenience, we refer to the clade composed by Leptopodomorpha and Terheteroptera hereafter as LCP. These three nodes have been widely accepted based on multiple lines of evidence (Schuh, 1979; Wheeler et al., 1993; Campbell et al., 1995; Sorensen

et al., 1995; Ouvrard et al., 2000; Xie et al., 2008; Schuh et al., 2009; Cryan and Urban, 2012; Spangenberg et al., 2013; Wang et al., 2016). In addition, their positions flank the problematic lineages in the tree of Hemiptera, including Fulgoromorpha, Cicadomorpha, Coleorrhyncha, Heteroptera and the basal lineages within Heteroptera.

In filtering the 3102 orthologous genes, we considered not only the monophyly of each control node but also the corresponding support value. The genes supporting the monophyly of Euhemiptera, Heteroptera and LCP were extracted using custom Python scripts and R. As a result, there are 176, 804 and 277 genes supporting the monophyly of Euhemiptera, Heteroptera and LCP respectively, with bootstrap values $\geq 50\%$. Three combinations, i.e., Euhemiptera together with LCP, Euhemiptera together with Heteroptera, and Euhemiptera alone, were used as control schemes to filter genes carrying phylogenetic signals. For the node combination of Euhemiptera and LCP, there were 45 genes supporting both clades with bootstrap values no lower than 50%. The clades Euhemiptera and Heteroptera shared 134 genes, which can simultaneously support them with bootstrap values no lower than 50%. The recovered intersection genes that have bootstrap values ≥ 60 , 70, 80 and 90% are listed in Table 2.

For all of the subsets based on different control schemes to gene filtering, phylogenetic trees were reconstructed using both concatenation and coalescent methods. For the ML analyses, node supports were estimated with 200 replicates using a rapid bootstrapping algorithm. Additionally, for the subsets which can simultaneously recover the Heteropterodea and Panheteroptera between ML analysis and coalescent

analysis, i.e., the 15-gene subset (Euhemiptera and LCP as controlled nodes, bootstrap values ≥ 70), 28-gene subset (Euhemiptera as controlled node, bootstrap values ≥ 90) and 35-gene subset (Euhemiptera and Heteroptera as controlled nodes, bootstrap values ≥ 80), Bayesian inference (BI) and MP analysis were conducted with MrBayes 3.2.6 using the Beagle library with CPU acceleration (Ronquist et al., 2012) and TNT (Goloboff and Catalano, 2016) respectively. For BI, a mixed amino-acid substitution model was used under a gamma distribution (+G) to account for among-site rate variation. Other parameters were set as follows: generations = 5 000 000, samplefreq = 100, printfreq = 100, nchains = 4, and those generations with standard deviation > 0.01 were discarded. For MP analysis, JK resampling was calculated with 2000 replicates. Traditional searches for trees were performed with 100 random replicates. The remaining parameter sets are the same as that for the concatenation analysis using 3102 genes.

Apomorphy mining of Heteroptera

The earliest fossils of Heteroptera can be dated back to 250 Ma (Shcherbakov, 2010). This means that specialized traits of the extant true bugs originated before then. To search for evidence at the molecular level, we applied a parsimony method to explore potential apomorphies of Heteroptera in the concatenated data set using PAUP*4.0a149 (Swofford, 2002) referring to the methods outlined by Wu et al. (2016). The two competing hypotheses of Coleorrhyncha relationships were taken into account to thoroughly mine the synapomorphies of Heteroptera. The topology inferred from the 3102-gene data set and 35-gene subset were used as the

Table 2
Recovery of Heteropterodea and Panheteroptera by the subsets based on node control

Controlled node(s)	Bootstrap value of controlled nodes	Number of genes	Heteropterodea		Panheteroptera	
			Concatenation analysis	Coalescent analysis	Concatenation analysis	Coalescent analysis
Euhemiptera and (Leptopodomorpha + Terheteroptera)	$\geq 50\%$	45	✓	✗	✗	✓
	$\geq 60\%$	29	✓	✗	✗	✓
	$\geq 70\%$	15*	✓	✓	✓	✓
	$\geq 80\%$	9	✓	✓	✗	✓
	$\geq 90\%$	3	✓	✓	✗	✗
Euhemiptera and Heteroptera	$\geq 50\%$	134	✓	✗	✗	✗
	$\geq 60\%$	93	✓	✓	✓	✗
	$\geq 70\%$	57	✓	✗	✓	✓
	$\geq 80\%$	35*	✓	✓	✓	✓
	$\geq 90\%$	17	✓	✓	✗	✗
Euhemiptera	$\geq 50\%$	176	✓	✗	✓	✗
	$\geq 60\%$	126	✓	✗	✓	✗
	$\geq 70\%$	80	✓	✗	✓	✗
	$\geq 80\%$	52	✓	✗	✓	✗
	$\geq 90\%$	28*	✓	✓	✓	✓

*The bold type is used to indicate the three subsets with strong signal.

Table 3
Fossils used for estimating divergence times and calibration point prior settings

Taxonomic group	Species	Locality	Geological age (Ma)	Mode of preservation	References
Thysanoptera: Liasothripidae	<i>Liasothrips crassipes</i>	Karatau-Galkino, Kazakhstan	Callovian–Oxfordian 166.1 ± 1.2 – 157.3 ± 1.0	Impression fossil	Shmakov (2008)
Thysanoptera: Thripidae	<i>Telthysrips libanicus</i>	Mdeyrii-Hammana, Casa Baabda, Lebanon	Barremian 129.4–125.0	Amber	Nel et al. (2010)
Hemiptera: Euhemiptera: Aviorrhynchidae	<i>Aviorrhyncha magnifica</i>	Terril No 7, Avion, France	Moscovian 315.2 ± 0.2 – 307.0 ± 0.1	Impression fossil	Nel et al. (2013)
Hemiptera: Aphidomorpha: Lutevanaphidae	<i>Lutevanaphis permiana</i>	F34, Les Vignasses, Lodève basin, France	Artinskian/Kungurian 290.1 ± 0.26 – 272.3 ± 0.5	Impression fossil	Szwedo et al. (2015)
Hemiptera: Fulgoromorpha: Fulgoridae	<i>Margaroptilon formosum</i>	Hondelage, Braunschweig, Germany	Early/Lower Toarcian 183.0–182.0	Impression fossil	Bode (1953)
Hemiptera: Cicadoidea: Tettigarctidae	<i>Leptoproshole lepida</i>	Bird's River (C-Dt II), South Africa	Carnian 237.0–227.0	Impression fossil	Riek (1976)
Hemiptera: Heteroptera: Enicocephalidae	<i>Enicocephalinus acragrimaldii</i>	Mdeyrii-Hammana, Casa Baabda, Lebanon	Barremian 129.4–125.0	Amber	Azar et al. (1999)
Hemiptera: Heteroptera: Schizopteridae	<i>Libanohypselosoma popovi</i>	Mdeyrii-Hammana, Casa Baabda, Lebanon	Barremian 129.4–125.0	Amber	Azar and Nel (2010)
Hemiptera: Heteroptera: Mesovelidae	<i>Gallomesovelia grioti</i>	River Dorches, northwest of Orbagnoux mine, France	Late/Upper Kimmeridgian 157.3 ± 1.0 – 152.1 ± 0.9	Compression fossil	Nel et al. (2014)
Hemiptera: Heteroptera: Velidae: Velinae	<i>Arcantivelia petraudi</i>	Font-de-Benon Quarry, Aisl-A level, France	Albian–Cenomanian boundary 105.3–99.6	Amber	Solórzano Kraemer and Perrichot (2014)
Hemiptera: Heteroptera: Naucoroidea: Triassocoridae	<i>Arlecoris lousi</i>	Vilsberg, France	Aegean 247.2–242.0	Impression fossil	Shcherbakov (2010)
Hemiptera: Heteroptera: Nepoidea: Belostomatidae	<i>Tarsabedus menkei</i>	0.6 km east Charmouth, UK	Sinemurian 199.3 ± 0.3 – 190.8 ± 1.0	Impression fossil	Popov et al. (1994)
Hemiptera: Heteroptera: Saldoidea: Archegocimicidae	<i>Britannicola senilis</i>	Apperley, UK	Hettangian 201.3 ± 0.2 – 199.3 ± 0.3	Impression fossil	Popov et al. (1994)
Hemiptera: Heteroptera: Cimicoidea: Vethanthocoridae	<i>Pumilanthocoris gracilis</i>	Daohugou, China	Callovian–Oxfordian 166.1 ± 1.2 – 157.3 ± 1.0	Compression fossil	Hou et al. (2012)
Hemiptera: Heteroptera: Miroidea: Miridae	<i>Scutellifer karatavicus</i>	Karatau-Mikhailovka	Callovian–Oxfordian 166.1 ± 1.2 – 157.3 ± 1.0	Impression fossil	Popov (1968)
Hemiptera: Heteroptera: Lygaeoidea: Pachymmeridiidae	<i>Ischnocoris bitoratus</i>	Hondelage, Braunschweig, Germany	Early/Lower Toarcian 183.0–182.0	Impression fossil	Bode (1953)

Table 4
Prior distributions of calibration points

Calibration point	Fossil species	Prior distributions and shapes (Ma)
Thysanoptera	<i>Liassothrips crassipes</i> <i>Tethystrips libanicus</i>	Normal, mean = 145.55, sigma = 10.2
Hemiptera	<i>Aviorrhyncha magnifica</i>	Normal, mean = 311.1, sigma = 4.0
Sternorrhyncha	<i>Lutevanaphis permiana</i>	Normal, mean = 283.5, sigma = 5.7
Fulgoromorpha	<i>Margaroptilon formosum</i>	Normal, mean = 182.7, sigma = 6.6
Cicadomorpha	<i>Leptoprobole lepida</i>	Normal, mean = 232.0, sigma = 4.9
Enicocephalomorpha	<i>Enicocephalinus acragrimaldii</i>	Normal, mean = 127.2, sigma = 2.8
Dipsocoromorpha	<i>Libanohypselosoma popovi</i>	Normal, mean = 127.2, sigma = 2.8
Gerromorpha	<i>Gallomesovelia grioti</i>	Normal, mean = 154.7, sigma = 5.2
Gerroidea	<i>Arcantivelia petraudi</i>	Normal, mean = 100.5, sigma = 3.4
Nepomorpha	<i>Arlecoris lousi</i>	Normal, mean = 244.6, sigma = 3.2
Nepoidea	<i>Tarsabedus menkei</i>	Normal, mean = 195.05, sigma = 2.9
Leptopodomorpha	<i>Britannicola senilis</i>	Normal, mean = 200.3, sigma = 4.0
Miroidea	<i>Scutellifer karatavicus</i>	Normal, mean = 163.5, sigma = 2.2
Cimiciformes	<i>Pumilanthocoris gracilis</i>	Normal, mean = 163.5, sigma = 2.2
Eutrichophora*	<i>Engerophorus nitidus</i>	Normal, mean = 182.7, sigma = 6.6

*Eutrichophora includes Coreoidea, Lygaeoidea and Pyrrhocoroidea.

input tree files, respectively. All taxa of Hemiptera were defined as ingroups. Parsimony character optimization was set to DELTRAN, which uses delayed transformation. After the log-file option was activated, sequence data were used to obtain a labelled tree with a complete list of apomorphies. Under the menu DescribeTrees, we selected the following options: the list of apomorphies, cladogram and label internal nodes. The results presented all possible apomorphies in the concatenated data set, which always include both real and ambiguous apomorphies. To avoid this scenario, the consistency index (CI) of each apomorphy was set to 1.0, which is the most stringent criterion, to filter for real amino acid apomorphies. A series of custom shell scripts were used to process the apomorphy list given by PAUP. There are a total of 1007 apomorphies of Heteroptera located in 553 orthologous genes.

Functions of orthologous genes

To classify and compare the functions of the genes carrying Heteroptera-specific characters, i.e., genes containing the amino acid apomorphies of Heteroptera, we annotated the functions for both the 3102 orthologous genes and the 553 orthologous genes using Blast2GO.v4.0 basic (Conesa et al., 2005; Conesa and Götz, 2008; Götz et al., 2008, 2011). The corresponding sequences of *Cimex lectularius* downloaded from OrthoDB were used to represent the 3102 genes and the 553 genes. The assemblies of the 3102 sequences and 553 sequences of *Cimex lectularius* were annotated separately. In Blast2GO, the BLAST similarity search was executed using blastp-fast against the non-redundant amino-acid sequence database (nr) of

the NCBI with an e-value cutoff of 1.0E-5. The remaining parameters were left at default settings. The gene ontology (GO) annotation results were plotted using the online service WEGO (Web Gene Ontology Annotation Plot) (Ye et al., 2006). The results in terms of molecular function, biological process and cellular components are provided as Fig. S1.

Divergence time estimation

Divergence times of deep nodes in Hemiptera–Heteroptera were inferred based on 15-gene, 28-gene and 35-gene subsets with BEAST.v2.4.1 (Bouckaert et al., 2014). BEAUti.v2.4.1 was used to generate the file used in the BEAST analysis. The JTT amino-acid substitution model was applied for all 15 genes due to the unavailability of the substitution model LG in BEAUti. The Birth–Death model of speciation (Stadler et al., 2013) and an uncorrelated log-normal relaxed clock (Drummond et al., 2006) were used. All other priors, except the calibration points described below, were left at default values in BEAST.

Sixteen fossil species from various lineages were used to calibrate 15 internal nodes to have comprehensive representatives and balanced calibrations among different lineages of the tree (Table 3). These fossils have been carefully checked according to the guidelines given by Parham et al. (2012), in particular their phylogenetic positions and geological age. Among the 15 calibrated nodes, each corresponds to one fossil species, except the Thysanoptera node, at which two fossil species were used. The oldest known fossil thrips, *Triassothrips virginicus*, was described from the Late Triassic of Virginia (228.0–208.5 Ma); it cannot be placed in any of the recent subclades and is considered to

represent an early, extinct lineage within Thysanoptera (Nel et al., 2012). Two species, *Liassothrips crassipes* (belonging to the extinct family Liassothripidae) and *Karataothrips jurassicus* (belonging to the extinct family Karataothripidae) (Shmakov, 2008; Nel et al., 2012), are known from the Late Jurassic of Kazakhstan (166.1–157.3 Ma); none of these can be unambiguously placed in either Terebrantia or Tubulifera. *Tethysthrips libanicus*, described from the Lower Cretaceous of Lebanon (129.4–125.0 Ma), can, however, be assigned to the extant family Thripidae (Nel et al., 2010). Therefore, for credibility, we used *L. crassipes* and *T. libanicus* to calibrate the node Thysanoptera. All calibrations were set as normal distributions, thus reflecting potential uncertainty in the fossil record and allowing variation of the posterior estimate in either direction (Ho and Phillips, 2009). The mean was set to the middle value of the Stage/Age in which the corresponding fossil was located, with the upper boundary of the previous interval and the lower boundary of the next interval as the 95% confidence interval bounds. Exact times corresponding to each geological stage were extracted from the International Chronostratigraphic Chart 2016 (Cohen et al., 2013). Prior distributions for them were tabulated for easy access (Table 4).

The analysis was run for 100 000 000 generations and was sampled every 100 generations using BEAST. Tracer v1.6 (<http://beast.bio.ed.ac.uk/Tracer>) was used to examine the posterior distribution of all parameters and their associated statistics, such as the effective sample size (ESS) and the 95% high posterior density (HPD) intervals. TreeAnnotator.v1.8 was used to summarize the set of post burn-in trees and their parameters and to produce a maximum clade credibility (MCC) chronogram showing mean divergence time estimates with 95% HPD intervals. All of the ESS values were above the recommended threshold of 200, indicating that the parameter space had been sufficiently sampled. An additional analysis was conducted to verify whether our prior settings had an influence on the results driving by data *per se*. This analysis was run under the prior settings without sequence data, with the remaining parameters the same as those in the normal analysis with sequence data.

Results

Data summary

In this study, we *de novo* sequenced transcriptomic data of 32 species, mainly in Heteroptera, with a particular focus on the infraorders Enicocephalomorpha, Dipsocoromorpha, Gerromorpha, Nepomorpha and Leptopodomorpha. The taxon sampling for Heteroptera was therefore complete at the infraorder level.

After aligning sequences, alignment masking and removing short alignments, 3102 orthologous genes were retained for further analyses. In the concatenated data set, the numbers of orthologous genes per taxon varied from 2148 (*Geisha distinctissima*) to 3102 (*Cimex lectularius*). The gene distribution occupied an average of ~85% of the concatenated data set (Fig. S2). The concatenated 3102-gene supermatrix comprises 769 646 aligned amino acid sites with a coverage of ~75% (all datasets used in this study are available at <http://www.dataopen.info/home/datafile/index/id/172>). The number of aligned amino acid sites per taxon varied from 575 171 (*Stenopirates* sp.) to 732 551 (*Acyrtosiphon pisum*) (Table 1).

Phylogenomic relationships based on the 3102 genes

Both the concatenation and the coalescent analyses produced a highly resolved phylogeny with 100% support values for most of the deep nodes of Hemiptera (Figs 2 and 3 and Fig. S3). All of the five suborders of Hemiptera and the seven infraorders of Heteroptera were recovered as monophyletic with 100% node support values. Sternorrhyncha was strongly supported as the sister group of Euhemiptera, and Auchenorrhyncha was recognized as monophyletic. Unexpectedly, Coleorrhyncha was placed as the sister group to Auchenorrhyncha, contradicting the widely accepted view of a monophyletic Heteropteroidea (= Coleorrhyncha + Heteroptera) based on multiple lines of evidence (Wheeler et al., 1993; Campbell et al., 1995; Sorensen et al., 1995; Ouvrard et al., 2000; Xie et al., 2008; Cryan and Urban, 2012; Spangenberg et al., 2013; Friedemann et al., 2014; Kuznetsova et al., 2015; Li et al., 2015).

At the base of Heteroptera, the topology of (Dipsocoromorpha + (Enicocephalomorpha + Gerromorpha)) received 100% support values. In addition, Terheteroptera and (Leptopodomorpha + Terheteroptera) within Heteroptera were suggested to be monophyletic. The monophyly of Panheteroptera (= Nepomorpha + Leptopodomorpha + Terheteroptera) was recovered with a 76% bootstrap value in the ML analysis and 100% support value in the MP analysis (Fig. 2 and Fig. S3). The phylogenetic tree generated by the coalescent analysis was highly congruent with that of the concatenation analysis except the position of Nepomorpha (Fig. 3), which clustered with a group composed of Dipsocoromorpha, Enicocephalomorpha and Gerromorpha with a 40% support value. No additional evidence, whether from molecular or morphological results, supports such a topology.

Phylogenetic relationships based on subsets

Because phylogenomic data may not correctly resolve all of the problematic nodes simultaneously,

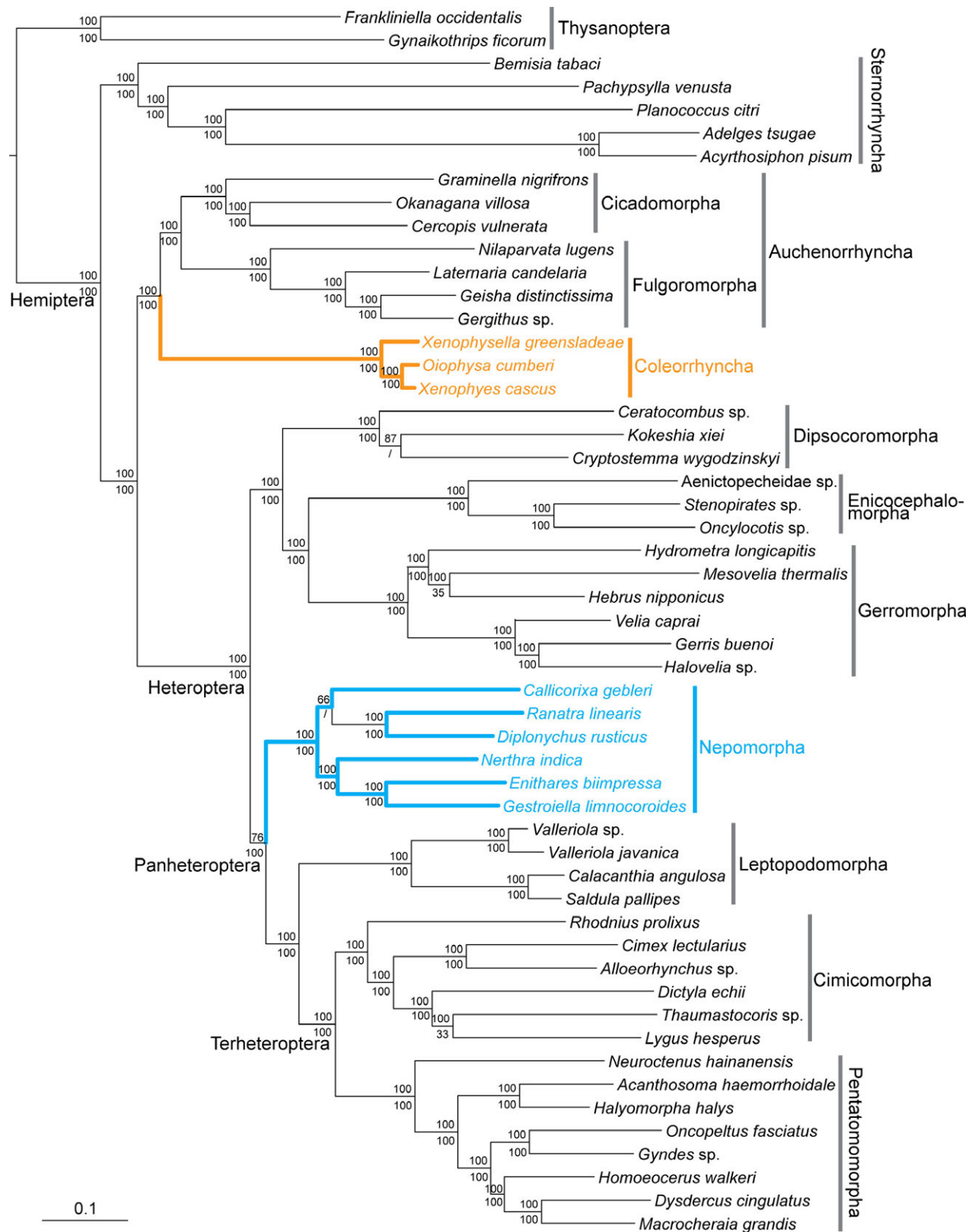


Fig. 2. Phylogenomic relationships of Hemiptera based on a 3102-gene concatenated data set. Numbers associated with each node indicate ML bootstrap values (upper) and MP support values (lower). The lengths of the branches follow the phylogram of the ML tree. Most of the major clades are labelled on the right of the tree. Scale bar denotes the number of substitutions per site. [Colour figure can be viewed at wileyonlinelibrary.com]

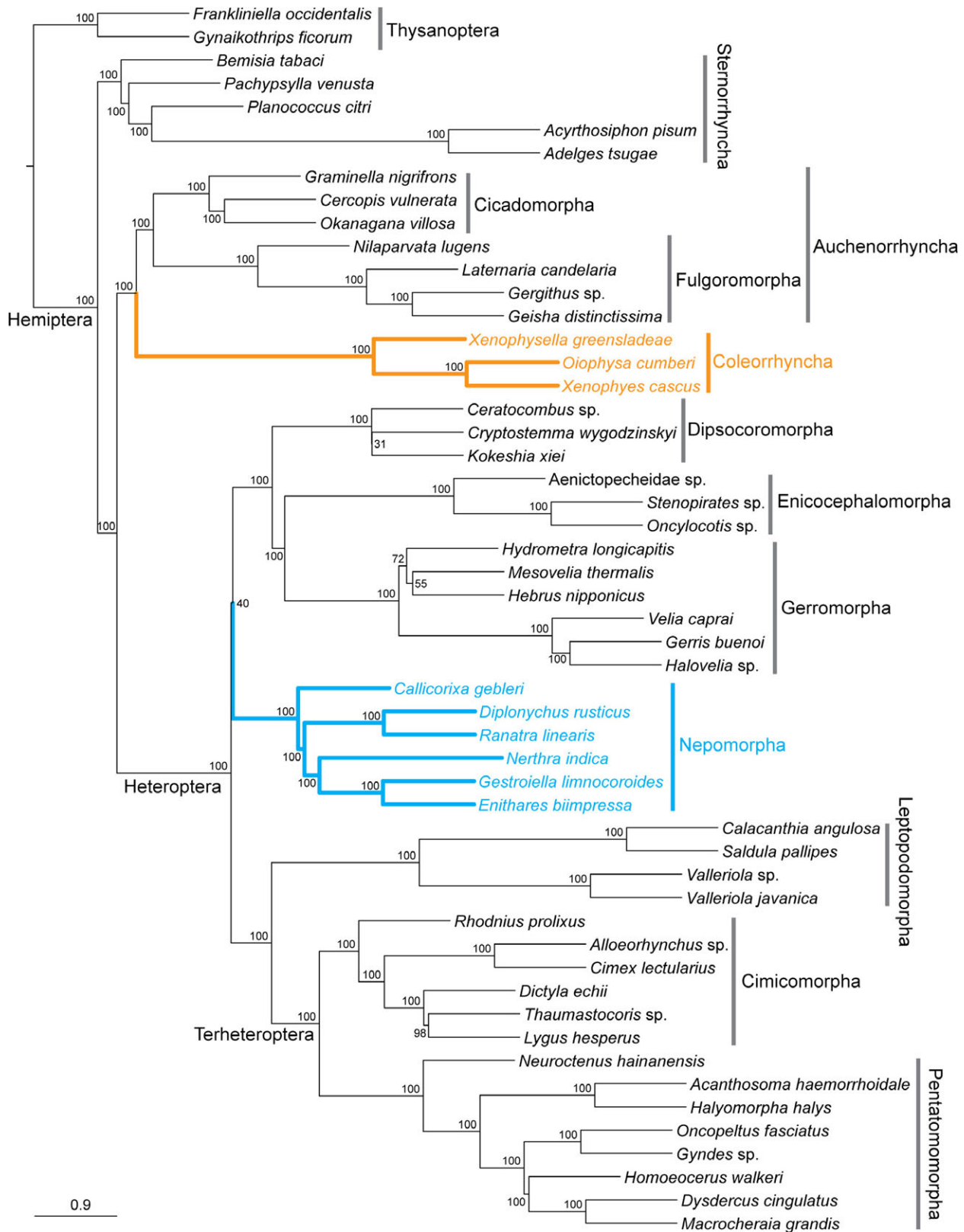


Fig. 3. Phylogenomic relationships of Hemiptera based on 3102 genes using a coalescent-based tree-inference method. Numbers associated with each node indicate Astral bootstrap values. Most of the major clades are labelled on the right of the tree. Scale bar indicates the number of substitutions per site. [Colour figure can be viewed at wileyonlinelibrary.com]

we used the optimization approach of binode control and uninode control to test whether the phylogenetic reconstruction of Hemiptera–Heteroptera can be resolved with congruence between different methods and analyses. The main focuses for the phylogenetics part of the present study were the positions of Coleorrhyncha and Nepomorpha in accordance with the unexpected or incongruent results based on the 3102 genes. Three widely accepted nodes, i.e., Euhemiptera, Heteroptera and LCP, which flank the problematic lineages mentioned above, were selected as the candidate nodes for control.

In a survey of the topology of the 3102 gene trees, 176, 804 and 277 genes supporting the clades Euhemiptera, Heteroptera and LCP with bootstrap values $\geq 50\%$ were identified, respectively. For the node combination of Euhemiptera and LCP, there are 45 genes supporting both clades with bootstrap values no lower than 50%. Meanwhile, the clades Euhemiptera and Heteroptera shared 134 genes which can simultaneously support both of them with bootstrap values no lower than 50%. Based on 45, 134 and 176 genes, we further generated a series of subsets corresponding to bootstrap values and thereafter reconstructed the phylogeny of Hemiptera–Heteroptera using both concatenation and coalescent methods (Table 2). For each scheme of binode or uninode control, there are five gene subsets corresponding to different bootstrap value thresholds.

The results of both concatenation and coalescent analyses using these subsets are listed in Table 2, and Fig. 4 and Figs S4–S38. All of the 15 ML analyses consistently support the monophyly of Heteropterodea (BS = 70–100%), and nine also support the monophyly of Panheteroptera (BS = 56–80%). The performance of coalescent analyses is relatively weak in recovering the monophyly of Heteropterodea or Panheteroptera, which received support in about half of the results for each. But importantly, in any of the three schemes of gene filtering by binode or uninode control, there is a subset that can recover the clades Heteropterodea and Panheteroptera by concatenation and coalescent analyses simultaneously (Fig. 4, Figs S12, S13, S22, S31 and S32). For the binode control based on the monophyly of Euhemiptera and LCP, the subset is a 15-gene subset; for the binode control using Euhemiptera and Heteroptera, the subset is a 28-gene subset; while for the uninode control using Euhemiptera alone, the subset is a 35-gene subset. Additional analyses of BI based on these three subsets also support the sister group of Coleorrhyncha and Heteroptera (PP = 100%) and that of Nepomorpha and LCP (PP = 95–100%) (Fig. 4 and Figs S33–S35). While in the MP analyses, although the monophyly of Heteropterodea and the sister group of Nepomorpha and LCP based on the 15-gene subset were not

recovered, the sister group of Nepomorpha and LCP obtained 99% and 91% support values based on the 28-gene and 35-gene subsets, respectively (Fig. 4 and Figs S36–S38). As for the remaining parts of the phylograms, the results based on the 15-, 28- and 35-gene subsets also supported the monophyly of Auchenorrhyncha (ML BS = 100, BI PP = 100, Astral BS = 52–92, JK = 98–100). The monophyly of a group composed by Dipsocoromorpha, Enicocephalomorpha and Gerromorpha was consistently recovered as well. The results of the latter two are in accordance with the result based on the 3102-gene data set.

Functional classifications of orthologous genes

Through apomorphy mining under the most stringent filter criterion, we located 1044 and 1255 potential apomorphies of amino acids shared by all extant true bugs according to the hypotheses (Coleorrhyncha + Heteroptera) and (Coleorrhyncha + Auchenorrhyncha), respectively. After manual checking and removal of those false positive apomorphies that are gap-dominant ($\geq 50\%$ taxa) or that are not exclusive to Heteroptera, 982 apomorphies of Heteroptera distributed in 542 orthologous genes were found for the hypothesis (Coleorrhyncha + Heteroptera), and 955 apomorphies of Heteroptera distributed in 533 orthologous genes were found for the hypothesis (Coleorrhyncha + Auchenorrhyncha). As the intersection result, 930 apomorphies located in 522 genes were shared according to both hypotheses. On the whole, 1007 apomorphies of Heteroptera in 553 genes were aggregated for subsequent analyses. Most of these genes (~85%) contain one or two amino acid apomorphies. However, in certain cases, the number of apomorphic amino acids in one orthologous gene reached up to 14 (zinc ion binding protein EOG091E01ZW in OrthoDB). The GO terms that have significant differences between the 3102 genes and 553 genes in terms of abundance, as calculated with WEGO using Pearson's chi-square test, are shown in Fig. 5.

To test whether stochastic sampling performed on the same number of genes would reach the same result, we randomly selected 553 genes and conducted the same functional annotations. This process was repeated three times. For two of the three replicates, there was only one GO term that showed significant promotion for the pseudo-sample of 553 genes, while no significant difference was detected for the third replicate.

Divergence time estimation

Deep divergence times of Hemiptera–Heteroptera based on the three subsets are consistent with each other. To make the description easier to follow, we

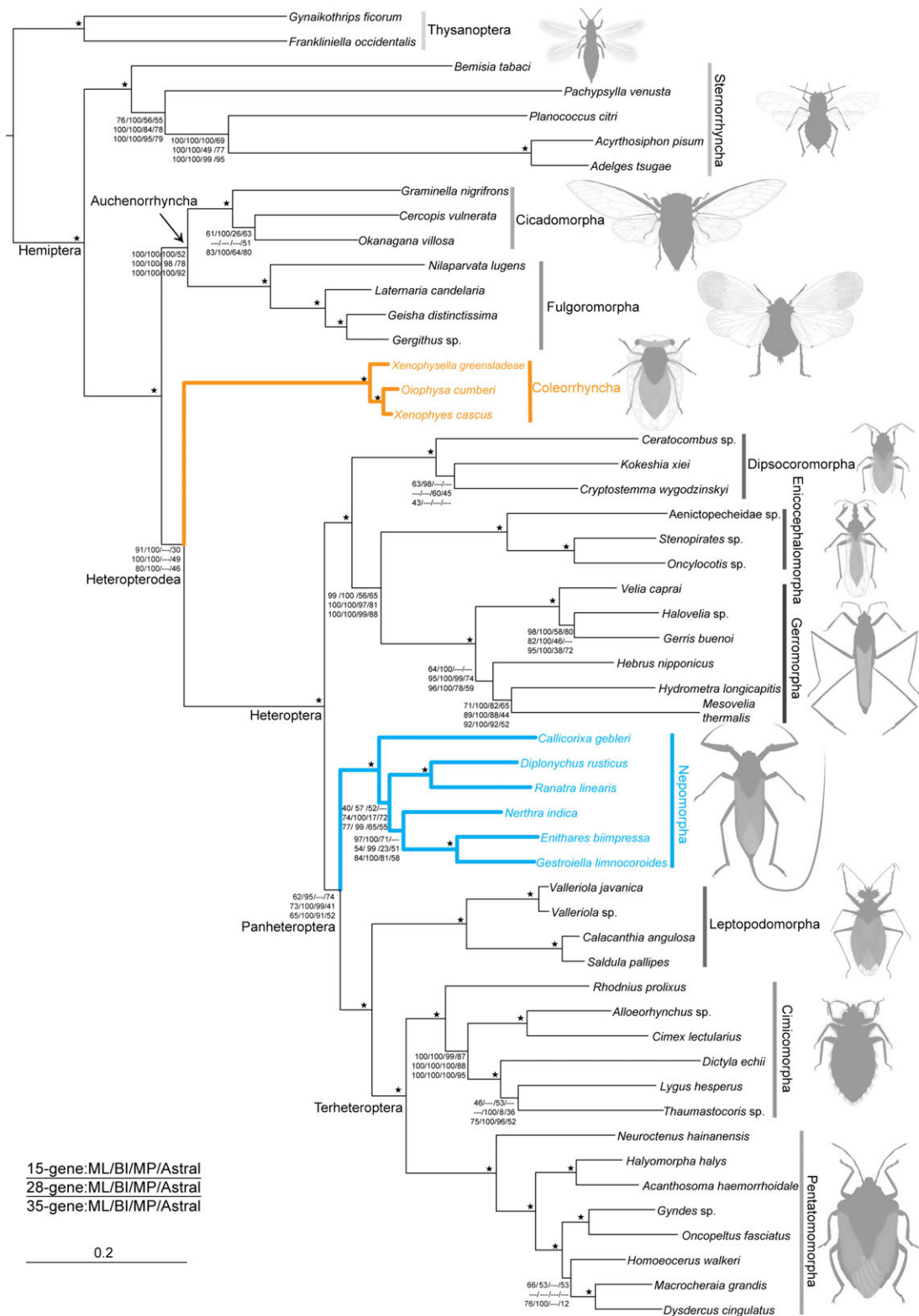


Fig. 4. Phylogenetic relationships based on subsets. Numbers associated with each node are (from left to right) ML bootstrap values, BI posterior probabilities, MP support values and Astral bootstrap values. The lengths of the branches follow the phylogram of the ML tree based on the 35-gene subset. Lack of an Astral bootstrap value indicates that the particular node was not recovered in the Astral analysis. Most of the major clades are labelled on the right of the tree. A star indicates that nodes all receive support values > 90%; otherwise, exact support values are indicated at each node. Scale bar indicates the number of substitutions per site. [Colour figure can be viewed at wileyonlinelibrary.com]

used the results inferred from the 35-gene subset. The MCC chronogram is displayed in Fig. 6 and Figs S39–S41, with a median node height and a 95% HPD interval for each node. Our results suggest that the origin of Heteropteroidea can be dated back to the Late Carboniferous (300 Ma, 95% HPD 290–309 Ma). The split between Coleorrhyncha and Heteroptera occurred 290 Ma (95% HPD 282–300 Ma) in the Early Permian. The divergence between Nepomorpha and LCP is estimated to have occurred 260 Ma (95% HPD 249–271 Ma) in the Middle and Late Permian, followed by a divergence between Leptopodomorpha and Terheteroptera 250 Ma (95% HPD 225–264 Ma) in the Early Triassic. The split between Cimicomorpha and Pentatomomorpha occurred 234 Ma (95% HPD 217–252 Ma) in the Carnian stage of the Triassic. In addition, for the basal clade of Heteroptera, the split between Dipsocoromorpha and (Enicocephalomorpha + Gerromorpha) occurred 228 Ma (95% HPD 202–247 Ma) in the Late Triassic, and the origin of Gerromorpha can be dated back to the Early Jurassic (197 Ma, 95% HPD 179–221 Ma).

Through comparison between the prior and posterior marginal distributions (File S2), we found that, for the calibration points, some of the posterior distributions of the age estimates were older, while some were younger than the prior distributions. This finding

indicates that our settings on the prior distributions did not influence the impact of the data set *per se*. That is, the 15 soft bound prior settings did not result in over-parameterization and dominate the posterior estimates.

Discussion

Despite previous efforts to clarify the phylogenetic history within Hemiptera, relationships between the suborders or infraorders remain uncertain, such as the position of Coleorrhyncha, the monophyly of Auchenorrhyncha, and the phylogenetic positions of Nepomorpha and Gerromorpha within Heteroptera. In this study, newly sequenced transcriptomic data of Fulgoromorpha, Coleorrhyncha, Enicocephalomorpha, Dipsocoromorpha and Leptopodomorpha enabled an in-depth phylogenetic analysis of Hemiptera–Heteroptera.

Phylogenetic position of Coleorrhyncha

The suborder Coleorrhyncha includes only one extant family, Peloridiidae. It is a relict group comprising ~36 described tiny (2–5 mm) recent species. Both the habitus and life style of peloridiids are cryptic and they live in wet moss in temperate and sub-Antarctic

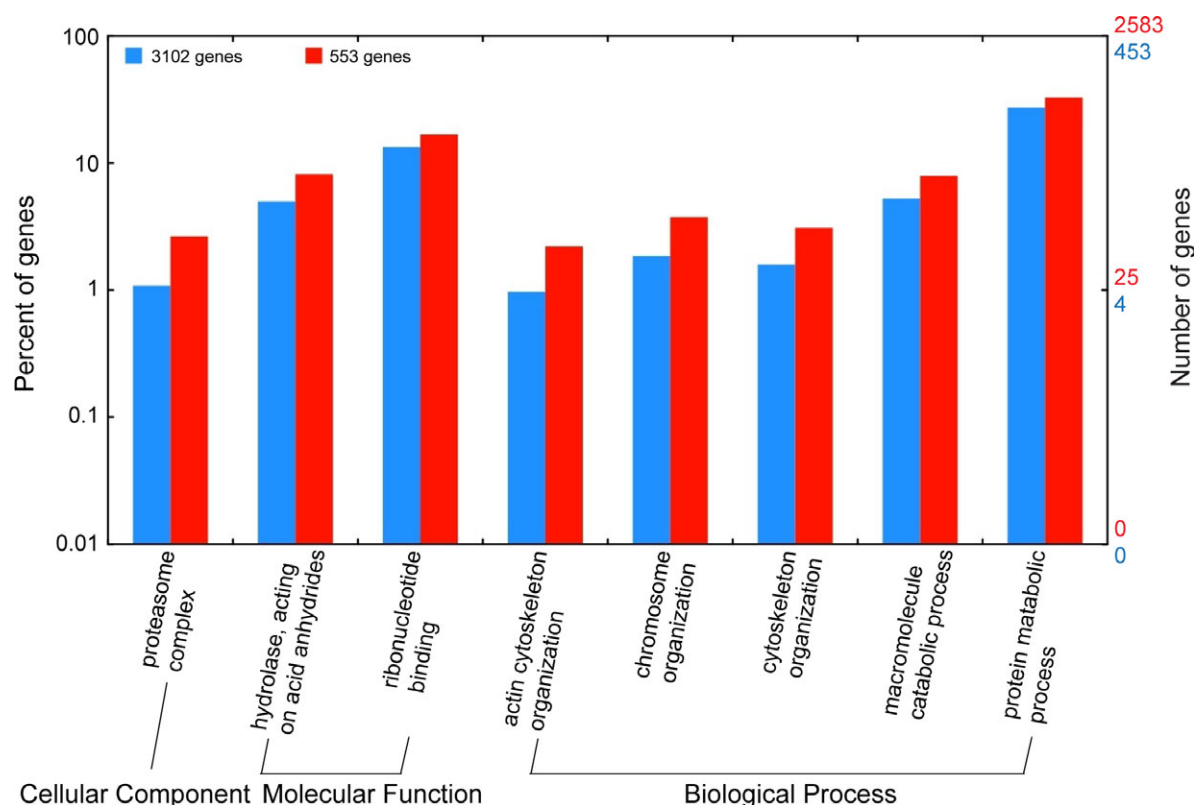


Fig. 5. Relative frequencies of functional gene groups that exhibit significant differences between the sets of 3102 (in blue) and 553 genes (in red). [Colour figure can be viewed at wileyonlinelibrary.com]

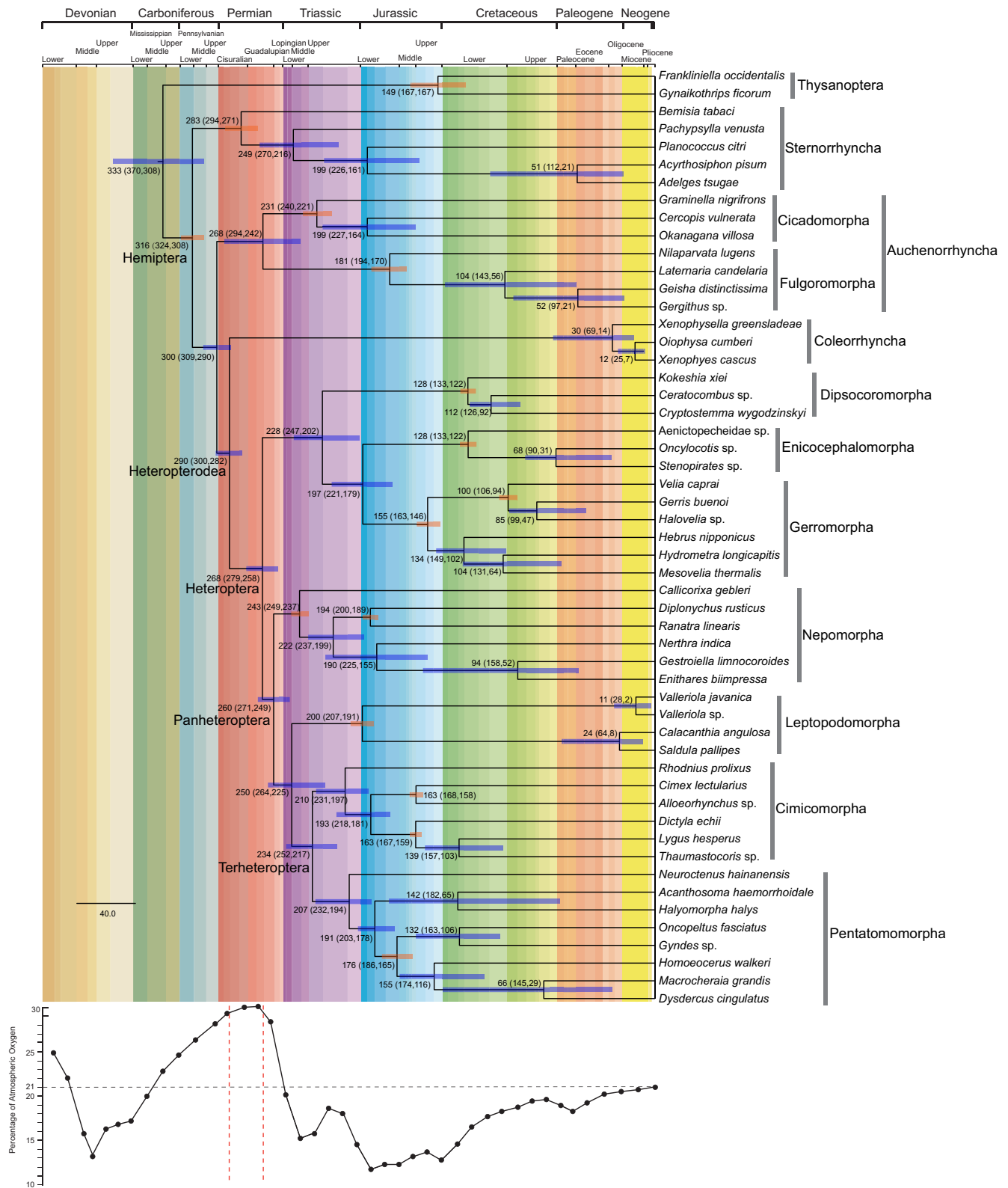


Fig. 6. Dated phylogenetic tree of Hemiptera based on the 35-gene subset. Horizontal blue bars indicate 95% credibility intervals. Red bars indicate calibration points. The diagram below the tree displays the percentage of atmospheric oxygen over time. Branch lengths are measured in millions of years. [Colour figure can be viewed at wileyonlinelibrary.com]

rainforests of South America, Australia, New Zealand and New Caledonia (Burckhardt et al., 2011). China (1962) briefly discussed the comparisons of wing and leg between Sternorrhyncha, Auchenorrhyncha, Coleorrhyncha (Peloridiidae) and Heteroptera. He suggested Peloridiidae are similar to Heteroptera in wing-coupling devices and legs, while similar to Auchenorrhyncha in veins and claws. The sister group relationship between Coleorrhyncha and Heteroptera was proposed by Schlee (1969) based on the following morphological characters (synapomorphies): reduced number of antennomeres; presence of intersegmental sclerites between particular antennal segments; “tracheal capture” (claval furrow is crossed by an anal vein which is united with cubitus in its distal section) (Wootton, 1965); abdomen dorsally flattened, forming a distinct connexivum laterally; anal tube without subdivisions; and presence of so-called basal plates at the proximal extremity of the male intromittent organ. Although Cobben (1978) provided a criticism of these characters, further synapomorphies potentially supporting a sister relationship between moss bugs and true bugs were suggested by subsequent authors, such as the similar wing-coupling structures (D’Urso, 1993), the presence of a distinct mandibular sulcus, the absence of clasping organs in the labial groove or the coiled accessory salivary ducts (Spangenberg et al., 2013). For the competitive hypothesis of (Coleorrhyncha + Auchenorrhyncha), only one line of evidence besides the phylogenomic one (Misof et al., 2014) has been raised, based on the presence of bacterial endosymbionts in bacteriomes (Müller, 1962; Buchner, 1965). However, the subsequent examination studies by Moran et al. (2005) showed that the presence of the bacterial endosymbiont is not that primary in some groups of Auchenorrhyncha, and thus weakened the significance of a bacterial endosymbiont to the phylogeny of host insects. Further phylogenetic studies showed that the bacterial endosymbionts can only support the monophyly of Auchenorrhyncha rather than Coleorrhyncha + Auchenorrhyncha (Moran et al., 2005; Kuechler et al., 2013; Santos-Garcia et al., 2014).

In the present study, although the position of Coleorrhyncha was suggested to be the sister group of Auchenorrhyncha using the results of both concatenation and coalescent methods based on the 3102 genes (Figs 2 and 3), the hypothesis of a sister relationship between Coleorrhyncha and Heteroptera based on the subsets should be given more attention. In fact, not all of the genes could contribute a signal to resolve the phylogenetic position of a certain node in a phylogenomic study (Salichos and Rokas, 2013; Chen et al., 2015). To decrease incongruence among genes, this study used the gene filtering method of binode and uninode control. According to the three different control schemes, no matter which one was employed,

congruent results between concatenation and coalescent methods can be used to recover the clade Heteropteroidea. Furthermore, this result is consistent with other evidence from molecular phylogenetics (Wheeler et al., 1993; Campbell et al., 1995; Sorensen et al., 1995; Ouvrard et al., 2000; Xie et al., 2008; Cryan and Urban, 2012; Li et al., 2015, 2017), morphological studies (Schlee, 1969; Spangenberg et al., 2013) and cytogenetic traits (Kuznetsova et al., 2015).

Aside from the congruent results reached by the concatenation and coalescent methods, and the congruence with the other lines of evidence, the signal quality of the three subsets with 15, 28 and 35 genes obtained by node control can also be justified by other arguments. First, no matter which scheme of node control was employed, there is a subset which can recover the monophyly of Heteropteroidea by concatenation and coalescent analyses congruently (Table 2). Second, the remaining part of the phylograms based on the subsets changed little in topology compared to the phylogram inferred from the concatenated 3102 genes, which means the three subsets carried enough phylogenetic signal to reconstruct the whole phylogeny of Hemiptera–Heteroptera. To test whether the subsets underestimated the actual numbers of potential signal genes, we then examined the heterogeneity between the 3102-gene trees. Under the filter conditions satisfying the monophylies of Auchenorrhyncha, Coleorrhyncha and Heteroptera, there are 27 and 20 candidate signal genes corresponding to the hypotheses of (Coleorrhyncha + Auchenorrhyncha) and (Coleorrhyncha + Heteroptera), respectively (bootstrap value $\geq 50\%$). Such results imply that, no matter which hypothesis about the phylogenetic position of Coleorrhyncha is true, the sizes of the potential signal data are at a similar level to those of the three subsets obtained by node control. In sum, the congruent results reached by different phylogenetic methods and comparative analyses showed that optimization of phylogenomic data based on node control can effectively extract signal genes and thus generate subsets for phylogenetic reconstruction.

If the 27 genes supporting (Coleorrhyncha + Auchenorrhyncha) and the 20 genes supporting (Coleorrhyncha + Heteroptera) are aggregated to generate another subset with 47 genes, the phylogenetic position of Coleorrhyncha as the sister group to Heteroptera can still be recovered by concatenation analysis (Figs S42 and S43). This result further indicated that the 20 genes supporting the monophyly of Heteropteroidea carry stronger signal than the 27 genes supporting the competitive hypothesis (Coleorrhyncha + Auchenorrhyncha). That is, a larger number of supporting genes for a competitive hypothesis does not necessarily mean more strength in evidence.

In nature, probably only a few genes contribute directly to speciation (Nosil and Schluter, 2011). Even

if more group-specific characters become fixed after speciation, the genomic similarity between two sister species could still be very high. When one or both of the species that have diverged undergo further speciation and/or anagenesis, highly specialized descendant lineages that additively accumulate a large number of autapomorphies might make the remaining ones appear more similar to one another due to symplesiomorphies. Considering the case of Coleorrhyncha, its potential sister group, Heteroptera, underwent a high level of specialization and diversification, while Coleorrhyncha retains many plesiomorphic characters shared with the early divergent lineages of Hemiptera, making it appear more similar to Auchenorrhyncha than to Heteroptera. As a consequence, simply increasing the amount of data does not guarantee correct topologies, which necessitates incongruence analyses and question-specific filtering of the data set.

It has long been recognized theoretically that genes qualified for resolving a particular phylogeny should meet certain criteria, especially a proper evolutionary rate, or in other words the balance between conservativeness and variability. However, due to limited available gene markers before the era of phylogenomics, there was little space for researchers to select genes other than mitogenomes, nuclear rDNAs and nuclear protein-coding genes. The idea of binode control is to select two widely accepted nodes flanking the problematic lineages. If a particular gene can simultaneously recover the monophyly of the controlled nodes, it is reasonable to suppose that it exhibits the right pattern of lineage sorting. This method thus filters for the most appropriate genes for phylogenetic inference of certain lineages at a practical level. Our study indicates the potential risk of using all available orthologous genes in a phylogenetic reconstruction without considering the incongruence between different genes and highlights the need to use question-specific methods.

Phylogenetic position of Auchenorrhyncha

Within Hemiptera, another hotly debated question is whether the suborder Auchenorrhyncha (including the extant superfamilies Fulgoroidea, Membracoidea, Cicadoidea and Cercopoidea) is a monophyletic group. Cryan and Urban (2012) comprehensively reviewed morphological and molecular studies concerning the phylogeny of Hemiptera before the year 2012. In addition, their study recovered the monophyly of Auchenorrhyncha based on a combination of nuclear and mitochondrial genes. Their work was the first one that had full taxon sampling and employed appropriate tree-inference methods. Since then, only a few phylogenetic studies based on mitochondrial genomes have explored the phylogeny of Hemiptera. Most of these studies considered Auchenorrhyncha as a paraphyletic

group (Cui et al., 2013; Song et al., 2016). However, Li et al. (2015, 2017) recovered Auchenorrhyncha as a monophyletic group.

In the present study, all of the phylogenetic analyses whether using the 3102-gene data set or using various subsets support the monophyly of Auchenorrhyncha. Such a result is in accordance with the evidence based on the combined mitochondrial and nuclear genes (Cryan and Urban, 2012) and morphological characters of, for example, the tymbal acoustic system (Ossiannilsson, 1949), forewing base (Yoshizawa and Saigusa, 2001) and wing-coupling structures (D'Urso, 2002).

Phylogenetic position of Nepomorpha

Concatenation analyses based on the 3102 genes and most of the subsets filtered by the optimization approach of binode and uninode control support the sister group relationship between Nepomorpha and LCP. This result is congruent with the results of phylogenetic reconstruction based on rDNAs and mitochondrial genomes (Wheeler et al., 1993; Wang et al., 2016). This hypothesis has also been supported by morphological characters such as the absence of arolia in adults, the distinctly hemielytral forewing, the presence of a forewing–body coupling mechanism of the “Druckknopf” system and the well-developed scutellar frena (Wheeler et al., 1993).

For the coalescent analyses, about half cannot reconstruct the monophyly of Panheteroptera. However, note that in all of the coalescent-based trees including the one based on the 3102-gene trees, the internode at the diversification of extant Heteroptera is very short. Such deep coalescence in the short internodes separating Nepomorpha, (Dipsocoromorpha + Enicocephalomorpha + Gerromorpha) and (Leptopodomorpha + Terheteroptera) can be a source of noise, generating real incongruence among gene trees. Incomplete lineage sorting is typically expected for shallow divergences, but discordance between gene trees can also remain after long time periods because lineage sorting only depends on the length of the internode and the effective population size and not on the depth of that internode (Degnan and Rosenberg, 2009; Edwards et al., 2016).

Origin of Heteroptera

Fifteen calibration points and soft bounds on priors were used to avoid the negative effects caused by selecting a single calibration point and excessively restricting priors (Ho and Phillips, 2009; Heled and Drummond, 2012). Our results suggest that the origin of Heteroptera occurred 290 Ma (95% HPD 282–300 Ma) in the Early Permian. The origin and

specialization of the common ancestor of true bugs occurred during a period from 290 to 268 Ma. Coincidentally, this span (268–290 Ma) is the time with the highest oxygen concentration in the atmosphere in Earth's history (Ward, 2006). In fact, the origin and/or the earliest diversification of several main clades in Insecta besides Heteroptera also occurred in the Permian, such as Orthoptera, Neuropterida, Coleoptera, Lepidoptera, and Diptera (Fig. S44) (Tong et al., 2015). The dual character of oxygen has long been recognized. On the one hand, aerobic organisms cannot live without oxygen, while on the other hand, high concentrations of atmospheric oxygen can promote the accumulation of harmful substances that increase mutation rates (Cooke et al., 2003). In the long evolutionary history of life, organisms have evolved mechanisms to repair damage caused by oxygen radicals, but some of this damage is irreversible and can be viewed as chronic, low-level, cumulative oxygen toxicity (Lenton, 2003).

Because the monophyly of (Coleorrhyncha + Auchenorrhyncha) was supported by the 3102-gene data set, we further tested the robustness of the origin time of Heteroptera based on the topology constructed by the concatenation analysis based on 3102 genes (Fig. 2 and Fig. S3). Another subset, which has equivalent gene numbers (35 genes) and can recover the same topology as that based on the 3102-gene data set using concatenation analysis, was generated for the test, and all of the parameter settings remained the same. The results of the test show that mean divergence times for most of the deep nodes fluctuated by a few million years (0–4 Ma) compared with those shown in Fig. 6 (Fig. S45). Although the origin of Heteroptera was pushed back approximately 15 Ma (274–305 Ma), it is still in the period with high and continuously increasing atmospheric oxygen.

Through functional annotations of the 3102 and 553 genes in the same window, the spectrum and possible functional biases of the genes carrying group-specific characters can provide in-depth features of genomic innovations. Our results show that the 553 genes covered most of the GO terms and ~70% of the third-level glossary terms of gene ontology. With regard to functional biases, there was increased abundance of genes carrying apomorphies of Heteroptera (Fig. 5). The functions of these genes are concentrated in macromolecule catabolic, protein metabolic and the other six items. These genes might play an important role in the biosynthesis of the true bugs' defensive secretions produced by the scent glands, which are the products of secondary metabolism of biomolecules. According to current knowledge of metabolic pathways in insects, some of the genes carrying apomorphies of Heteroptera indeed participate in the synthesis of biomolecules (Morgan, 2010).

The broad spectrum of genes carrying Heteroptera-specific characters and the increased abundance of those that are functional in metabolic processes affecting biomolecules probably indicate two phases in the origin of true bugs. At first, the common ancestor of true bugs probably experienced large-scale non-adaptive evolution at the genome level. After the emergence of the phenotype possessing scent glands, more genes affecting the trait of the biochemical pathways and morphological structures responsible for producing the defensive secretion of these insects were further intensified by adaptive evolution. These latter fixations further strengthened the traits associated with being “stinky”. Moreover, this model of “non-adaptive evolution + adaptive evolution” may provide a new hypothesis for the seemingly sudden emergence of a new complex trait in evolution, and can be more explanatory than the hypothesis of “hopeful monster”. These findings shed light on the probable impact of global changes in atmospheric oxygen on the genomic innovation of the common ancestor of true bugs.

Early diversification of Heteroptera

According to our phylogenetic results, the earliest divergence of extant Heteroptera was a split into the clades Dipsocoromorpha + Enicocephalomorpha + Gerromorpha and Panheteroptera. Based on our divergence time estimates, the origin of Gerromorpha can be dated back to the Norian stage of the Triassic (214 Ma, 95% HPD 193–240 Ma). The Nepomorpha clade separated from LCP around the Capitanian stage of the Permian (261 Ma, 95% HPD 253–271 Ma), which is 47 million years earlier than the origin of Gerromorpha. Gerromorpha has generally been considered as more basal than Nepomorpha, and members of both infraorders occur in similar habitats that are associated with water, so one might expect that they have a similar chance of fossilization (Damgaard, 2008). However, almost all of the oldest definitive fossils of Heteroptera belong to Nepomorpha (middle Triassic, ~250 Ma), while the oldest fossils of Gerromorpha are from the late Jurassic (~150 Ma). Our results offer an explanation for this apparent contradiction.

The split between Leptopodomorpha and Terheteroptera and the divergence between Cimicomorpha and Pentatomomorpha inferred from the present study are both consistent with earlier results based on mitogenomic protein-coding genes and nuclear rDNAs (Wang et al., 2016). Diversification of the superfamilies within Cimicomorpha and Pentatomomorpha were found to have occurred during the Jurassic, which is consistent with the hypothesized late Jurassic origin of angiosperms based on molecular dating (Smith et al., 2010). The origin and early radiation of angiosperms

not only supplied substantial nutrition resources for phytophagous insects but also offered a high diversity of heterogeneous niches for these insects.

Conclusions

Our phylogenomic studies of Hemiptera support that the phylogenetic position of Heteroptera is as the sister group to Coleorrhyncha. All five suborders of Hemiptera and the seven infraorders of Heteroptera were recovered as monophyletic. The relationship within Heteroptera is ((Dipsocoromorpha + (Enicocephalomorpha + Gerromorpha)) + (Nepomorpha + (Leptopodomorpha + (Cimicomorpha + Pentatomomorpha)))). The results of divergence time estimation showed that the common ancestor of true bugs originated 290–268 Ma in the Permian, when atmospheric oxygen was the richest in Earth's history. This timespan also witnessed the specialization or rapid diversification of other major clades of insects. During this period, at least 553 genes had accumulated Heteroptera-specific amino acids, and some of them are incorporated in the biosynthesis of scent substrates, which probably suggest non-adaptive evolution before adaptive evolution in the origin of true bugs.

Acknowledgements

We thank George W. Gibbs (Victoria University, New Zealand) for collecting two rare species of Peloridiidae and preparing the corresponding RNAlater preserved samples. We thank Gang Wang and Xiao-Guang Liu (Nankai University, China) for providing useful suggestions. We thank K. Jun Tong and Nathan Lo (University of Sydney, Australia) for personal communications. This work was supported by the projects of the National Natural Science Foundation of China (grant numbers: 31222051, 31472024, 31572242). The authors declare no conflicts of interest.

References

- Azar, D., Nel, A., 2010. The earliest fossil schizopterid bug (Insecta: Heteroptera) in the Lower Cretaceous amber of Lebanon. *Ann. Soc. Entomol. Fr.* 46, 193–197.
- Azar, D., Fleck, G., Nel, A., Solignac, M., 1999. A new enicocephalid bug, *Enicocephalinus acragrimaldii* gen. nov., sp. nov., from the Lower Cretaceous amber of Lebanon (Insecta, Heteroptera, Enicocephalidae). *Est. Mus. Cienc. Nat. de Álava* 14, 217–230.
- Betancur-R, R., Naylor, G., Ortí, G., 2014. Conserved genes, sampling error, and phylogenomic inference. *Syst. Biol.* 63, 257–262.
- Bode, A., 1953. Die Insektenfauna des Ostniedersächsischen Oberen Lias. *Palaeontographica* 103, 1–375.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, e1003537.
- Bourgoin, T., 1988. A new interpretation of the homologies in the Hemiptera male genitalia, illustrated by the Tettigometridae (Hemiptera, Fulgoromorpha). *Proceedings of the Sixth International Auchenorrhyncha Meeting, University of Turin Press, Turin, Italy*, pp. 113–120.
- Bourgoin, T., 1993. Female genitalia in Hemiptera Fulgoromorpha, morphological and phylogenetic data. *Ann. Soc. Entomol. Fr.* 29, 225–244.
- Buchner, P., 1965. *Endosymbiosis of Animals with Plant Microorganisms*. Interscience Publishers, New York, NY.
- Burckhardt, D., Bochud, E., Damgaard, J., Gibbs, G.W., Hartung, V., Larivière, M., Wyniger, D., Zürcher, I., 2011. A review of the moss bug genus *Xenophyes* (Hemiptera: Coleorrhyncha: Peloridiidae) from New Zealand: systematics and biogeography. *Zootaxa* 2923, 1–26.
- Campbell, B.C., Steffen-Campbell, J.D., Gill, R.J., 1994. Evolutionary origin of whiteflies (Hemiptera: Sternorrhyncha: Aleyrodidae) inferred from 18S rDNA sequences. *Insect Mol. Biol.* 3, 73–89.
- Campbell, B.C., Steffen-Campbell, J.D., Sorensen, J.T., Gill, R.J., 1995. Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. *Syst. Entomol.* 20, 175–194.
- Chen, M., Liang, D., Zhang, P., 2015. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. *Syst. Biol.* 64, 1104–1120.
- China, W.E., 1962. South American Peloridiidae (Hemiptera-Homoptera: Coleorrhyncha). *Trans. R. Entomol. Soc. Lond.* 114, 131–161.
- Cobben, R.H., 1978. *Evolutionary Trends in Heteroptera*. Pt. II. Mouthpart-structures and Feeding Strategies. Agricultural University, Wageningen.
- Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.X., 2013. The ICS international chronostratigraphic chart. *Episodes* 36, 199–204.
- Collins, T.M., Fedrigo, O., Naylor, G.J.P., 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. *Syst. Biol.* 54, 493–500.
- Conesa, A., Götz, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008, 1–13.
- Conesa, A., Götz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676.
- Cooke, M.S., Evans, M.D., Dizdaroglu, M., Lunec, J., 2003. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* 17, 1195–1214.
- Cryan, J.R., Urban, J.M., 2012. Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Syst. Entomol.* 37, 7–21.
- Cui, Y., Xie, Q., Hua, J.-M., Dang, K., Zhou, J.-F., Liu, X.-G., Wang, G., Yu, X., Bu, W.-J., 2013. Phylogenomics of Hemiptera (Insecta: Paraneoptera) based on mitochondrial genomes. *Syst. Entomol.* 38, 233–245.
- Damgaard, J., 2008. Evolution of the semi-aquatic bugs (Hemiptera: Heteroptera: Gerromorpha) with a re-interpretation of the fossil record. *Acta. Entomol. Mus. Natl. Pragae* 48, 251–268.
- De Noya, B.A., Díaz-Bello, Z., Colmenares, C., Ruiz-Guevara, R., Mauriello, L., Zavala-Jaspe, R., Suarez, J.A., Abate, T., Naranjo, L., Paiva, M., Rivas, L., Castro, J., Márques, J., Mendoza, I., Acquatella, H., Torres, J., Noya, O., 2010. Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. *J. Infect. Dis.* 20, 1308–1315.
- Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference, and the multispecies coalescent. *Trends Ecol. Evol.* 24, 332–340.
- Dell'Ampio, E., Meusemann, K., Szucsich, N.U., Peters, R.S., Meyer, B., Borner, J., Petersen, M., Aberer, A.J., Stamatakis, A.,

- Walz, M.G., Minh, B.Q., von Haeseler, A., Ebersberger, I., Pass, G., Misof, B., 2014. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol. Biol. Evol.* 31, 239–249.
- Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375.
- Doggett, S.L., Dwyer, D.E., Penas, P.F., Russell, R.C., 2012. Bed bugs: clinical relevance and control options. *Clin. Microbiol. Rev.* 25, 164–192.
- von Dohlen, C.D., Moran, N.A., 1995. Molecular phylogeny of the Homoptera: a paraphyletic taxon. *J. Mol. Evol.* 41, 211–223.
- Drummond, A., Ho, S.Y., Phillips, M., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4, e88.
- D'Urso, V., 1993. The wing-coupling apparatus in *Peloidium hammomiorum* Bredding, 1897 (Insecta, Rhynchota). *Spixiana* 16, 133–139.
- D'Urso, V., 2002. The wing-coupling apparatus of Hemiptera Auchenorrhyncha: structure, function, and systematic value. *Denisia* 176, 401–410.
- Ebersberger, I., Strauss, S., von Haeseler, A., 2009. HaMStR: profile hidden Markov model based search for orthologs in ESTs. *BMC Evol. Biol.* 9, 157.
- Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Mol. Phylogenet. Evol.* 94, 447–462.
- Footitt, R., Adler, P., 2009. *Insect Biodiversity: Science and Society*. Blackwell Publishing, London.
- Friedemann, K., Spangenberg, R., Yoshizawa, K., Beutel, R.G., 2014. Evolution of attachment structures in the highly diverse Acercaria (Hexapoda). *Cladistics* 30, 170–201.
- Garipey, T.D., Haye, T., Fraser, H., Zhang, J., 2014. Occurrence, genetic diversity, and potential pathways of entry of *Halymorpha halyis* in newly invaded areas of Canada and Switzerland. *J. Pest. Sci.* 87, 17–28.
- Goloboff, P.A., Catalano, S.A., 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* 32, 221–238.
- Goodchild, A.J.P., 1966. Evolution of the alimentary canal of the Hemiptera. *Biol. Rev.* 41, 97–140.
- Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435.
- Götz, S., Arnold, R., Sebastián-León, P., Martín-Rodríguez, S., Tischler, P., Jehl, M.A., Dopazo, J., Rattei, T., Conesa, A., 2011. B2G-FAR, a species centered GO annotation repository. *Bioinformatics* 27, 919–924.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.
- Heled, J., Drummond, A.J., 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst. Biol.* 61, 138–149.
- Henry, T.J., 2009. Biodiversity of Heteroptera. In: Footitt, R.G., Adler, P.H. (Eds.), *Insect Biodiversity: Science and Society*. Blackwell Publishing, Oxford, pp. 223–264.
- Ho, S.Y., Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* 58, 367–380.
- Hou, W.-J., Yao, Y.-Z., Zhang, W.-T., Ren, D., 2012. The earliest fossil flower bugs (Heteroptera: Cimicomorpha: Cimicoidea: Vethanthocoridae) from the Middle Jurassic of Inner Mongolia, China. *Eur. J. Entomol.* 109, 281–288.
- Hu, D.L., Chan, B., Bush, J.W.M., 2003. The hydrodynamics of water strider locomotion. *Nature* 424, 663–666.
- Irisarri, I., Meyer, A., 2016. The identification of the closest living relative(s) of Tetrapods: phylogenomic lessons for resolving short ancient internodes. *Syst. Biol.* 65, 1057–1075.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., Wägele, J.W., Misof, B., 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front. Zool.* 7, 10.
- Kuechler, S.M., Gibbs, G., Burckhardt, D., Dettner, K., Hartung, V., 2013. Diversity of bacterial endosymbionts and bacteria–host co-evolution in Gondwanan relict moss bugs (Hemiptera: Coleorrhyncha: Peloridiidae). *Environ. Microbiol.* 15, 2031–2042.
- Kuznetsova, V.G., Grozeva, S.M., Hartung, V., Anokhin, B.A., 2015. First evidence for (TTAGG)_n telomeric sequence and sex chromosome post-reduction in Coleorrhyncha (Insecta, Hemiptera). *Comp. Cytogenet.* 9, 523–532.
- Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29, 1695–1701.
- Lartillot, N., Brinkmann, H., Philippe, H., 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site heterogeneous model. *BMC Evol. Biol.* 7, S4.
- Lemmon, E.M., Lemmon, R., 2013. High-throughput genomic data in systematics and phylogenetics. *Annu. Rev. Ecol. Evol. Syst.* 44, 99.
- Lenton, T.M., 2003. The coupled evolution of life and atmospheric oxygen. In: Rothschild, L.J., Lister, A.M. (Eds.), *Evolution on Planet Earth: The Impact of the Physical Environment*. Academic Press, London, pp. 35–54.
- Li, H., Shao, R.-F., Song, N., Song, F., Jiang, P., Li, Zh.-H., Cai, W.-Zh., 2015. Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences. *Sci. Rep.* 5, 8527.
- Li, H., Leavengood, J.M. Jr, Chapman, E.G., Burckhardt, D., Song, F., Jiang, P., Liu, J., Zhou, X., Cai, W., 2017. Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc. R. Soc. B* 284, 20171223.
- Meusemann, K., von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kück, P., Ebersberger, I., Walz, M., Pass, G., Breuers, S., Achter, V., von Haeseler, A., Burmester, T., Hadrys, H., Wägele, J.W., Misof, B., 2010. A phylogenomic approach to resolve the arthropod tree of life. *Mol. Biol. Evol.* 27, 2451–2464.
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. Astral: genome-scale coalescent species tree estimation. *Bioinformatics* 30, i541–i548.
- Misof, B., Misof, K., 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst. Biol.* 58, 21–34.
- Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermini, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walz, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L.,

- Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M. and Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767.
- Moran, N.A., Tran, P., Gerardo, N.M., 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Appl. Environ. Microbiol.* 71, 8802–8810.
- Morgan, E.D., 2010. *Biosynthesis in Insects* (Advanced Edition). The Royal Society of Chemistry Publishing, London.
- Müller, H.J., 1962. Neuere Vorstellungen über Verbreitung und Phylogenie der Endosymbiosen der Zikaden. *Zoomorphology* 51, 190–210.
- Nel, P., Peñalver, E., Azar, D., Hodebert, G., Nel, A., 2010. Modern thrips families Thripidae and Phlaeothripidae in Early Cretaceous amber (Insecta: Thysanoptera). *Ann. Soc. Entomol. Fr. (n.s.)* 46, 154–163.
- Nel, P., Azar, D., Prokop, J., Roques, P., Hodebert, G., Nel, A., 2012. From Carboniferous to recent: wing venation enlightens evolution of thysanopteran lineage. *J. Syst. Palaeontol.* 10, 385–399.
- Nel, A., Roques, P., Nel, P., Prokin, A.A., Bourgoin, T., Prokop, J., Szewo, J., Azar, D., Desutter-Grandcolas, L., Wappler, T., Garrouste, R., Coty, D., Huang, D.-Y., Engel, M.S., Kirejtshuk, A.G., 2013. The earliest known holometabolous insects. *Nature* 503, 257–261.
- Nel, A., Nel, P., Krieg-Jacquier, R., Pouillon, J.M., Garrouste, R., 2014. Exceptionally preserved insect fossils in the Late Jurassic lagoon of Orbagnoux (Rhône Valley, France). *PeerJ* 2, 1–16.
- Nosil, P., Schluter, D., 2011. The genes underlying the process of speciation. *Trends Ecol. Evol.* 26, 160–167.
- Ossiannilsson, F., 1949. Insect drummers. A study on the morphology and function of the sound-producing organ of Swedish Homoptera Auchenorrhyncha. *Opusc. Entom. Suppl.* 10, 1–146.
- Ouvrard, D., Campbell, B.C., Bourgoin, T., Chan, K.L., 2000. 18S rDNA secondary structure and phylogenetic position of Peloridiidae (Insecta, Hemiptera). *Mol. Phylogenet. Evol.* 16, 403–417.
- Parham, J.F., Donoghue, P., Bell, C.J., Calway, T.D., Head, J.J., Holroyd, P.A., Inoue, J.G., Irmis, R.B., Joyce, W.G., Ksepka, D.T., Patané, J.S., Smith, N.D., Tarver, J.E., van Tuinen, M., Yang, Z., Angielczyk, K.D., Greenwood, J.M., Hipsley, C.A., Jacobs, L., Makovicky, P.J., Müller, J., Smith, K.T., Theodor, J.M., Warnock, R.C., Benton, M.J., 2012. Best practices for justifying fossil calibrations. *Syst. Biol.* 61, 346–359.
- Perlea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J., 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19, 651–652.
- Philippe, H., Roure, B., 2011. Difficult phylogenetic questions? More data, maybe; better methods, certainly. *BMC Biol.* 9, 91.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houlston, E., Quéinnec, E., Da Silva, C., Wincker, P., Le Guyader, H., Leys, S., Jackson, D.J., Schreiber, F., Erpenbeck, D., Morgenstern, B., Wörheide, G., Manuel, M., 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* 19, 706–712.
- Popov, Y.A., 1968. True bugs of the Jurassic Karatau fauna (Heteroptera). In: Rohdendorf, B.B. (Ed.), *Jurassic Insects of Karatau, Acad. Sci. Moscow, Section of General Biology*. Publ. House “Nauka”, Moscow, pp. 99–113.
- Popov, Y.A., Dolling, W.R., Whalley, P.E.S., 1994. British Upper Triassic and Lower Jurassic Heteroptera and Coleorrhyncha (Insecta: Hemiptera). *Genus* 5, 307–347.
- Regier, J.C., Shultz, J.W., Ganley, A.R.D., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J.E., Cummings, M.P., Martin, J.W., Cunningham, C.W., 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Syst. Biol.* 57, 920–938.
- Riek, E.F., 1976. A new collection of insects from the Upper Triassic of South Africa. *Ann. Natal Mus.* 22, 791–820.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
- Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., Douzery, E.J.P., 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. *Mol. Biol. Evol.* 30, 2134–2144.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542.
- Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497, 327–331.
- Santos-Garcia, D., Latorre, A., Moya, A., Gibbs, G., Hartung, V., Dettner, K., Kuechler, S.M., Silva, F.J., 2014. Small but powerful, the primary endosymbiont of moss bugs, *Candidatus Evansia muelleri*, holds a reduced genome with large biosynthetic capabilities. *Genome Biol. Evol.* 6, 1875–1893.
- Schlee, D., 1969. Morphologie und Symbiose; ihre Beweiskraft für die Verwandtschaftsbeziehungen der Coleorrhyncha (Insecta, Hemiptera). *Stuttg. Beitr. Naturk.* 210, 1–27.
- Schuh, R.T., 1979. Evolutionary trends in Heteroptera. Part II. Mouthpart-structures and feeding strategies. *Syst. Zool.* 28, 653–656.
- Schuh, R.T., Slater, J.A., 1995. *True Bugs of the World (Hemiptera: Heteroptera) Classification and Natural History*. Cornell University Press, Ithaca, NY.
- Schuh, R.T., Weirauch, C., Wheeler, W.C., 2009. Phylogenetic relationships within the Cimicomorpha (Hemiptera: Heteroptera): a total-evidence analysis. *Syst. Entomol.* 34, 15–48.
- Shcherbakov, D.E., 2010. The earliest true bugs and aphids from the Middle Triassic of France (Hemiptera). *Russ. Entomol. J.* 19, 179–182.
- Shcherbakov, D.E., Popov, Y.A., 2002. Superorder Cimicidea Laicharting, 1781. Order Hemiptera Linné, 1758. The bugs, cicadas, plantlice, scale insects, etc. In: Rasnitsyn, A.P., Quicke, D.L.J. (Eds.), *History of Insects*. Kluwer Academic, Dordrecht, pp. 143–157.
- Shmakov, A.S., 2008. The Jurassic thrips *Liassothrips crassipes* (Martynov, 1927) and its taxonomic position in the order Thysanoptera (Insecta). *Paleontol. J.* 42, 47–52.
- Simmons, M.P., Carr, T.G., O'Neill, K., 2004. Relative character-state space, amount of potential phylogenetic information, and heterogeneity of nucleotide and amino acid characters. *Mol. Phylogenet. Evol.* 32, 913–926.
- Slater, J.A., 1982. Hemiptera. In: Parker, S. (Ed.), *Synopsis and Classification of Living Organisms*, Vol. 2. McGraw-Hill, New York, pp. 417–447.
- Smith, S.A., Beaulieu, J.M., Donoghue, M.J., 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl Acad. Sci. USA* 107, 5897–5902.
- Solórzano Kraemer, M.M., Perrichot, V., 2014. Taxonomic names, in fossil water striders in Cretaceous French amber (Heteroptera: Gerromorpha: Mesoveliidae and Veliidae). *Syst. Entomol.* 39, 590–605.
- Song, N., Liang, A.-P., Bu, C.-P., 2012. A molecular phylogeny of Hemiptera inferred from mitochondrial genome sequences. *PLoS ONE* 7, e48778.
- Song, N., An, Sh.-H., Yin, X.-M., Cai, W.-Zh., Li, H., 2016. Application of RNA-seq for mitogenome reconstruction, and reconsideration of long-branch artifacts in Hemiptera phylogeny. *Sci. Rep.* 6, 33465.
- Sorensen, J.T., Campbell, B.C., Gill, R.J., Steffen-Campbell, J.D., 1995. Non-monophyly of Auchenorrhyncha (“Homoptera”), based upon 18S rDNA phylogeny: eco-evolutionary and cladistic implications within pre-Heteropterodea Hemiptera (*s.l.*) and a proposal for new monophyletic suborders. *Pan-Pac. Entomol.* 71, 31–60.

- Spangenberg, R., Wipler, B., Friedemann, K., Pohl, H., Weirauch, C., Hartung, V., Beutel, R.G., 2013. The cephalic morphology of the Gondwanan key taxon *Hackeriella* (Coleorrhyncha, Hemiptera). *Arthropod Struct. Dev.* 42, 315–337.
- Stadler, T., Kühnert, D., Bonhoeffer, S., Drummond, A.J., 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl Acad. Sci. USA* 110, 228–233.
- Stamatakis, A., 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Štys, P., Kerzhner, I., 1975. The rank and nomenclature of higher taxa in recent Heteroptera. *Acta Entomol. Bohemoslov.* 72, 65–79.
- Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.
- Szwedo, J., Lapeyrie, J., Nel, A., 2015. Taxonomic names, in Rooting down the aphid's tree – the oldest record of the Aphidomorpha lineage from Palaeozoic (Insecta: Hemiptera). *Syst. Entomol.* 40, 207–213.
- Tong, J.K., Duchene, S., Ho, S.Y.W., Lo, N., 2015. Comment on “Phylogenomics resolves the timing and pattern of insect evolution”. *Science* 349, 487.
- Wang, Y.-H., Cui, Y., Rédei, D., Bañar, P., Xie, Q., Štys, P., Damgaard, J., Chen, P.-P., Yi, W.-B., Wang, Y., Dang, K., Li, Ch.-R., Bu, W.-J., 2016. Phylogenetic divergences of the true bugs (Insecta: Hemiptera: Heteroptera), with emphasis on the aquatic lineages: the last piece of the aquatic insect jigsaw originated in the Late Permian/Early Triassic. *Cladistics* 32, 390–405.
- Ward, P.D., 2006. Out of Thin Air: Dinosaurs, Birds, and Earth's Ancient Atmosphere. Joseph Henry Press, Washington, DC.
- Weirauch, C., Štys, P., 2014. Litter bugs exposed: phylogenetic relationships of Diposcoromorpha (Hemiptera: Heteroptera) based on molecular data. *Insect Syst. Evol.* 45, 351–370.
- Wheeler, W.C., Schuh, R.T., Bang, R., 1993. Cladistic relationships among higher groups of Heteroptera: congruence between morphological and molecular data sets. *Entomol. Scand.* 24, 121–137.
- Wootton, R.J., 1965. Evidence for tracheal capture in early Heteroptera. *Proceedings of the 12th International Congress of Entomology*, pp. 65–67.
- Wu, H.-Y., Wang, Y.-H., Xie, Q., Ke, Y.-L., Bu, W.-J., 2016. Molecular classification based on apomorphic amino acids (Arthropoda, Hexapoda): integrative taxonomy in the era of phylogenomics. *Sci. Rep.* 6, 28308.
- Xie, Q., Tian, Y., Zheng, L.-Y., Bu, W.-J., 2008. 18S rRNA hyperelongation and the phylogeny of Euhemiptera (Insecta: Hemiptera). *Mol. Phylogenet. Evol.* 47, 463–471.
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., Wang, J., 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297.
- Yoshizawa, K., Saigusa, T., 2001. Phylogenetic analysis of paraneopteran orders (Insecta: Neoptera) based on forewing base structure, with comments on monophyly of Auchenorrhyncha (Hemiptera). *Syst. Entomol.* 26, 1–13.
- Zhang, N., Zeng, L., Shan, H., Ma, H., 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* 195, 923–937.
- Zrzavý, J., 1992. Evolution of antennae and historical ecology of the hemipteran insects (Paraneoptera). *Acta Entomol. Bohemoslov.* 89, 77–86.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Detailed GO annotation of 3102 genes and 553 genes.

Figure S2. Gene distribution for each taxon relative to the 3102 orthologous genes.

Figure S3. MP tree based on the 3102-gene concatenated data set. Numbers associated with each node indicate MP support values. Scale bar denotes the number of substitutions per site.

Figure S4. ML tree based on the concatenation of 45 genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S5. Coalescent-based phylogenetic tree based on 45 genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S6. ML tree based on the concatenation of 134 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S7. Coalescent-based phylogenetic tree based on 134 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S8. ML tree based on the concatenation of 176 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S9. Coalescent-based phylogenetic tree based on 176 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S10. ML tree based on the concatenation of 29 genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S11. Coalescent-based phylogenetic tree based on 29 genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S12. ML tree based on the concatenation of 15 genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S13. Coalescent-based phylogenetic tree based on 15 genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S14. ML tree based on the concatenation of nine genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S15. Coalescent-based phylogenetic tree based on nine genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S16. ML tree based on the concatenation of three genes using Euhemiptera and LCP as controlled nodes.. Scale bar indicates the number of substitutions per site.

Figure S17. Coalescent-based phylogenetic tree based on three genes using Euhemiptera and LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S18. ML tree based on the concatenation of 93 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S19. Coalescent-based phylogenetic tree based on 93 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S20. ML tree based on the concatenation of 57 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S21. Coalescent-based phylogenetic tree based on 57 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S22. Coalescent-based phylogenetic tree based on 35 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S23. ML tree based on the concatenation of 17 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S24. Coalescent-based phylogenetic tree based on 17 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S25. ML tree based on the concatenation of 126 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S26. Coalescent-based phylogenetic tree based on 126 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S27. ML tree based on the concatenation of 80 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S28. Coalescent-based phylogenetic tree based on 80 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S29. ML tree based on the concatenation of 52 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S30. Coalescent-based phylogenetic tree based on 52 genes using Euhemiptera as controlled

node. Scale bar indicates the number of substitutions per site.

Figure S31. ML tree based on the concatenation of 28 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S32. Coalescent-based phylogenetic tree based on 28 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S33. Bayesian inference based on the concatenation of 15 genes using Euhemiptera plus LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S34. Bayesian inference based on the concatenation of 28 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S35. Bayesian inference based on the concatenation of 35 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S36. MP tree based on the concatenation of 15 genes using Euhemiptera plus LCP as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S37. MP tree based on the concatenation of 28 genes using Euhemiptera as controlled node. Scale bar indicates the number of substitutions per site.

Figure S38. MP tree based on the concatenation of 35 genes using Euhemiptera and Heteroptera as controlled nodes. Scale bar indicates the number of substitutions per site.

Figure S39. Dated phylogenetic tree of Hemiptera based on 15-gene subset. Branch lengths are measured in millions of years.

Figure S40. Dated phylogenetic tree of Hemiptera based on 28-gene subset. Branch lengths are measured in millions of years.

Figure S41. Dated phylogenetic tree of Hemiptera based on 35-gene subset. Branch lengths are measured in millions of years.

Figure S42. ML tree based on the concatenation of the top 47 genes, which is the sum of genes supporting the two hypotheses of Coleorrhyncha. Scale bar indicates the number of substitutions per site.

Figure S43. Coalescent-based phylogenetic tree based on the top 47 genes, which is the sum of genes supporting the two hypotheses of Coleorrhyncha. Scale bar indicates the number of substitutions per site.

Figure S44. Lineage increase of insects during the past 576 million years based on the chronogram of Tong et al. (2015). The green lines indicate the period with atmospheric oxygen >28%.

Figure S45. Dated phylogenetic tree of Hemiptera based on a similar 35-gene subset which can recover the same topology as the ML tree based on the concatenation of 3102 genes.

File S1. Protocols of how to establish core-orthologous genes and local databases of reference species.

File S2. Prior and posterior marginal distributions. Grey indicates the posterior density distributions and blue denotes the prior density distributions.